Philipps Universität Marburg

Penalized likelihood based tests for regime switching in autoregressive models

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultäten

der Philipps-Universität Marburg

vorgelegt von

Florian Ketterer

Dipl. Math. oec.

aus Achern

Erstgutachter: Prof. Dr. Hajo Holzmann

Zweitgutachter: Prof. Dr. Norbert Henze

Eingereicht: 08.04.2011

Tag der mündlichen Prüfung: 16.06.2011

Contents

Introduction

A large variety of time series models, such as linear autoregressive or autoregressive conditional heteroscedastic (ARCH) models, are used to analyze the dynamic behavior of economic or financial variables. Since time series often undergo changes in their behavior over time, associated with events such as financial crises, such constant parameter time series models might be inadequate for describing the data.

The *Markov-switching model* of Hamilton (1989) is one of the most popular regime switching models in the literature. This model involves multiple structures that characterize the time series' behavior in different regimes. While the original Markov-switching model mainly focuses on the mean behavior of the time series, incorporating the switching mechansim into linear autoregressive models, Cai (1994) and Hamilton and Susmel (1994) studied various ARCH models with Markov switching, incorporating the switching mechanism into conditional variance models. An important feature of the Markov-switching model is that the switching mechanism is controlled by an unobservable state variable that follows a first-order Markov chain. The determination of the number of states in the hidden Markov chain is a task of major importance. In this thesis we are mainly concerned with the basic methodological issue to test for regime switching, i.e. we are testing for the existence of at least two states, in various Markov-switching autoregressive models. Since, under the hypothesis, parameters of the full model are not identifiable the asymptotic distribution of the corresponding likelihood ratio test is highly nonstandard. This problem already arises in the closely related problem of testing for homogeneity in two-component mixtures. To overcome this non-identifiability problem Chen, Chen and Kalbfleisch (2001) developed a penalized likelihood ratio test which admits a simple asymptotic distribution. Additional difficulties arise if the Markov dependence structure is incorporated into the test statistic. Therefore, Cho and White (2007) propose a quasi likelihood ratio test (QLRT) for regime switching in general autoregressive models which neglects the dependence structure of the hidden Markov chain under the alternative. We extend their approach using penalized likelihood based tests in order to obtain tractable asymptotic distributions of several test statistics.

In Chapter 1 we introduce Markov-switching autoregressive and closely related models and discuss the methodology we use.

The modified likelihood ratio test introduced by Chen, Chen and Kalbfleisch (2001) is well established for testing for homogeneity in finite mixture models. In Chapter 2 we extend this test to Markov-switching autoregressive models with a univariate switching parameter which fulfill some regularity conditions. These regularity conditions are satisfied by (*i*) linear switching autoregressive models with switching variance and t- or normal innovations, linear switching autoregressive models with a univariate switching autoregressive parameter and t- or normal innovations, linear switching autoregressive models with switching intercept and t-innovations and

(*ii*) switching ARCH models with switching intercept in the ARCH part with t- or normal innovations.

We show that the asymptotic distribution of the modified (quasi) likelihood ratio test under the hypothesis is given by a mixture of a point mass at zero and a χ_1^2 distribution with equal weights. Finally, we introduce a closely related test, called EM-test, which admits the same asymptotic distribution as the modified (quasi) likelihood ratio test.

For applications, the linear switching autoregressive model with switching intercept and normal innovations is very important, cf. Hamilton (2008). It is desirable to develop feasible methods for testing for homogeneity in this model. Studying asymptotic properties of test statistics which are based on the (penalized) likelihood becomes very challenging since $\sigma \frac{\partial^2 f(x;\mu,\sigma)}{\partial^2 \mu} = \frac{\partial f(x;\mu,\sigma)}{\partial \sigma}$ holds for the normal distribution. Here, $f(x;\mu,\sigma)$ denotes the density of a normal distribution with mean μ and standard deviation $\sigma > 0$. This problem already arises when testing for homogeneity in homoscedastic normal mixture models, for which Chen and Li (2009) investigated a method for testing. In Chapter 3 we extend their approach to linear switching autoregressive models where the intercept switches according to the underlying regime. We show that the asymptotic distribution of the corresponding test statistic under the hypothesis is a simple function of a shifted χ_1^2 and a $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ distribution. We also propose a test based on fixed proportions under the alternative. Under the hypothesis, the asymptotic distribution of the corresponding test statistic is a function of a χ_1^2 and a $\frac{1}{2}\chi_0^2+\frac{1}{2}\chi_1^2$ distribution. We apply the methods developed in Chapter 2 and 3 to the series of seasonally adjusted quarterly U.S. GNP data from 1947(1)–2002(3) and find a regime switch in the volatility of the growth rate. Dividing the series in two subseries 1947(1)–1984(1) and 1984(2)–2002(3), we cannot find clear evidence of a regime switch in the intercept of a linear autoregressive model in these subseries.

In Chapter 4 we are concerned with testing for homogeneity in a linear switching autoregressive model where the intercept as well as the scale parameter of the normally distributed innovations are allowed to switch. To this end, we extend the EM-test introduced by Chen and Li (2009) for testing for homogeneity in a normal mixture model with possibly distinct means and variances under the alternative. We show that the asymptotic distribution of our test statistic under the hypothesis is given by a χ_2^2 distribution. Since the EM-test admits the same asymptotic distribution if $\alpha = 1/2$ is fixed under the alternative we also propose a test based on fixed proportion $\alpha = 1/2$ under the alternative. Therefore, feasible methods for testing for homogeneity in a model which is used (in a slightly different version) for modeling stock returns, see Bhar and Hamori (2004), have been found. We apply our methods to the series of monthly log returns of the IBM stock. We find evidence of two states: Regime 1 with lower mean level and higher variance and regime 2 with higher mean level and lower variance.

Acknowledgments

First of all, I am very grateful to my supervisor Prof. Dr. Hajo Holzmann, for giving me the opportunity to carry out this research, for constant encouragement and for being open for discussions and questions.

Moreover, I also wish to thank Prof. Dr. Norbert Henze for taking the Korreferat.

Special thanks go to Dipl. Math. oec. Daniel Hohmann and Anne-Christin Zimmer for reading parts of this thesis and providing me many valuable comments.

I am also indebted to Dr. Jörn Dannemann for some helpful hints and inspiring discussions.

I would like to thank my office-mates in Karlsruhe and Marburg for helpful discussions and frequent cheering up.

The friends that have helped me are too numerous to mention here. But, most importantly, I want to thank Dr. Mario Hörig, Anika Beer, Christian Obloh, Nadine and Thorsten Banzhaf.

I gratefully acknowledge financial support from the DFG, grant HO 3260/3-1 and from the Landesstiftung Baden-Württemberg, 'Juniorprofessorenprogramm'.

Last and important thanks go to my family for various forms of support, above all from my parents Erich and Gerlinde as well as from my sisters Carola and Iris.

In loving memory of my uncle Rolf Braun (1967-2010)

1 Markov-switching autoregressive and related models

Switching autoregressive models are parametric time series models in which parameters are allowed to take on different values in each of some fixed number of regimes. A stochastic process assumed to have generated the regime shifts is included as part of the model. For *Markov-switching autoregressive models* we usually assume that the regime shifts occur according to a Markov chain with finite state space. In general the process generating the regime shifts cannot be observed. However, in *self-exciting threshold models* we assume that regime shifts are triggered by the level of an observable variable in relation to an unobserved threshold. In this thesis we focus on *Markov-switching autoregressive models* which are a good choice for modeling nonlinear time series if there is no a priori knowledge about deterministic events, such as the excess of a threshold value leading to a regime switch. Instead, regime switches will occur rather suddenly. For modeling e.g. exchange rates, however, *self-exciting threshold models* seem to be an appropriate choice since there will be an intervention by the government when the exchange rate exhibits certain upper or lower thresholds.

In this chapter we introduce *Markov-switching autoregressive models* which belong to the class of latent variable models. Latent variable models can be used to model complex data structures which are given by the observations by introducing latent variables. Sometimes these unobservable variables have a theoretical justification or are motivated by some desirable interpretation such as different volatility states in stock returns. Models which are closely related to *Markov-switching autoregressive models*, including *hidden Markov models* and *finite mixture models*, will also be treated in this chapter. These models have in common that the hidden variables form a discrete time stochastic process on some finite set $\mathcal{M} = \{1, \ldots, m\}$, say.

1.1 Finite mixture models

Finite mixture models are convenient for describing populations with unobserved heterogeneity. Many monographs deal with all kinds of properties appearing in the literature, including identifiability and parameter estimation. For an overview see McLachlan and Peel (2000), Frühwirth-Schnatter (2006) or Titterington, Smith and Makov (1985). A recent survey article about mixture models is given by Seidel (2010).

A famous example concerning finite mixture models is due to Hosmer (1973). According to the International Halibut Commission of Seattle, Washington, the length distribution of halibut of a given age is well approximated by a mixture of two normal distributions corresponding to the length distributions of the male and female subpopulation: Denoting the observations by X_k and the membership to one of the populations by S_k, this formalizes to $P(X_k \leq x|S_k = 1) = \Phi\big((x - \mu_1)/\sigma_1\big)$, $P(X_k \leq x|S_k = 2) = \Phi\big((x - \mu_2)/\sigma_2\big)$ and $P(S_k = 2) = 1 - P(S_k = 1) = \alpha$, where $\Phi(\cdot)$ is the cdf of a standard normal variate. Assuming that $(S_k)_k$ and $(X_k)_k$ are two independent sequences (but not independent of each other) leads to a univariate two component mixture model of two normal distributions with distribution function

$$\begin{aligned} G(x) = P(X_k \leq x) =& P(S_k = 1)P(X_k \leq x|S_k = 1) + P(S_k = 2)P(X_k \leq x|S_k = 2) \\ =& (1 - \alpha)\Phi\big((x - \mu_1)/\sigma_1\big) + \alpha\Phi\big((x - \mu_2)/\sigma_2\big) \end{aligned}$$

with parameter $(\alpha, \mu_1, \mu_2, \sigma_1, \sigma_2)$.

In general, an m-component mixture distribution reads

$$G(x) = \alpha_1 F_1(x) + \ldots + \alpha_m F_m(x), \tag{1.1.1}$$

where $\alpha_j \geq 0$, $\sum_{j=1}^m \alpha_j = 1$, and F_j specifies the distribution ot the jth component. As in the example above, the latent variable here represents the unobservable membership to one of the components, and (1.1.1) arises from

$$G(x) = P(X_k \leq x) = P(S_k = 1)P(X_k \leq x|S_k = 1) + \ldots + P(S_k = m)P(X_k \leq x|S_k = m),$$

where $S_k \sim Mult(1; \boldsymbol{\alpha})$ are i.i.d. multinomial random variables on $\{1, \ldots, m\}$. If not stated otherwise, in this thesis we assume that the state dependent distributions $P(X_k \leq x|S_k = j) = F_j(x)$, $j = 1, \ldots, m$, belong to the same parametric family indexed by a parameter $\vartheta \in \Theta \subset \mathbb{R}^l$, $l \geq 1$, i.e. $F_j = F_{\vartheta_j}$. Hence, the parameter of interest is

$$\boldsymbol{\omega} = (\alpha_1, \ldots, \alpha_{m-1}, \vartheta_1, \ldots, \vartheta_m).$$

Identifiability of finite mixtures

In general, a parametric family of distributions indexed by a finite dimensional parameter $\boldsymbol{\omega}$ which is defined over a sample space $\mathcal{X}$ is said to be identifiable if any two parameters $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ induce the same probability law on $\mathcal{X}$ if and only if $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ coincide. In terms of the corresponding probability densities $p(x; \boldsymbol{\omega})$ and $p(x; \boldsymbol{\omega}')$ w.r.t. to some σ-finite measure ν on $\mathcal{X}$ this means that the parameters $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ coincide if the densities are identical for ν-almost all $x \in \mathcal{X}$.

Clearly, the family of univariate normal distributions indexed by $\boldsymbol{\omega} = (\mu, \sigma^2)$ is identifiable, whereas for (finite) mixtures of probability distributions the issue of identifiability is much more involved, see e.g. McLachlan and Peel (2000). For finite mixture models, assuming that the F_j's belong to the same parametric family, identifiability was studied exhaustively and is established for e.g. finite mixtures of Poisson distributions (Feller, 1943), of normal and gamma distributions (Teicher, 1963), of multivariate normal distributions (Yakowitz and Spragins, 1968) and of binomial distributions $Bin(n_0, p)$, $0 < p < 1$ and n_0 fixed, provided that $n_0 > 2m - 1$ (Teicher, 1963). Sometimes it might be helpful to use an equivalent characterization due to Yakowitz and Spragins (1968). They show that the class of finite mixtures of distributions is identifiable if and only if the underlying parametric family is linearly independent over the field of real numbers $\mathbb{R}$.

Dealing with identifiability in mixture distributions it is convenient to distinguish the following three types of non-identifiability:

(i) Non-identifiability due to invariance to relabeling the components of a mixture,

(ii) non-identifiability due to potential overfitting and

(iii) generic non-identifiability.

There are many attempts for ruling out the first type of non-identifiability in the literature. Standard approaches to overcome this problem are changing the nomenclature to equivalence classes w.r.t. label switching (Leroux, 1992b) or ordering the parameters of the distributions of the components (e.g. $\vartheta_1 < \ldots < \vartheta_m$, where we use the lexicographical order if $\Theta \subset \mathbb{R}^l$, $l \geq 2$). While the consideration of equivalence classes is uncomfortable from a practical point of view, ordering constraints may not be desirable in some applications, see e.g. Frühwirth-Schnatter (2006). Especially if the distributions of two components are close to each other in some sense ordering the parameters can have a significant influence on statistical inference. To overcome the second type of non-identifiability we have to assume that the number of components m is known, i.e. we have $\alpha_i > 0$ and $F_i \neq F_j$ for $1 \leq i \neq j \leq m$.

It is worth mentioning that there are also some examples of finite mixture distributions which remain unidentifiable even if we rule out any of the first two non-identifiability issues such as finite location-scale mixtures of triangular distributions, see Holzmann, Munk and Gneiting (2006, Ex. 6).

Parameter estimation in finite mixtures

Let $X_1, \ldots, X_n$ be an i.i.d. sample from a finite mixture model. The parameter of interest is

$$\boldsymbol{\omega} = (\alpha_1, \ldots, \alpha_{m-1}, \vartheta_1, \ldots, \vartheta_m),$$

which can be estimated via different methods. Classical approaches are method of moments, Bayesian Estimation or Maximum Likelihood (see Frühwirth-Schnatter, 2006, for a short overview about these methods). For the maximum likelihood estimator (MLE) we have to compute the argument maximizing

$$\tilde{l}_n(\boldsymbol{\omega}) = \tilde{l}_n(\boldsymbol{\omega}; X_1, \dots, X_n) = \sum_{k=1}^{n} \log\Big(\sum_{i=1}^{m} \alpha_i f_{\vartheta_i}(X_k)\Big),$$

where f_{ϑ_i} denotes the density corresponding to the conditional distribution F_{ϑ_i} w.r.t. some σ-finite measure ν on $\mathcal{X}$.

Since the MLE often cannot be calculated explicitly it has to be assigned numerically. Two ways to compute the MLE are Newton type algorithms or the expectation maximization (EM) algorithm, introduced by Dempster, Laird and Rubin (1977). It is designed for models with incomplete information. In case of mixture models we can regard $S_1, \dots, S_n$, indicating the group membership of the observations $X_1, \dots, X_n$, as missing.

Maximum likelihood estimation has been used for univariate mixtures of two homoscedastic normal distributions, i.e. $\sigma_1^2 = \sigma_2^2$, regarding the variance as a structural parameter, as early as Rao (1948) while testing for homogeneity in this model has been an open problem until Chen and Li (2009).

Determining the number of components in finite mixture models

If there is no a priori knowledge about the number m of components in a mixture model, testing for this number is an important but difficult issue which has not been completely resolved, yet. Testing for the number of components is known to be difficult since it often involves inference for an overfitted mixture model where the true number of components is less than the number of components in the fitted mixture model, as e.g. in case of the likelihood ratio test (LRT). Parameter estimation in this case represents a non-regular problem with the true parameter lying in a non-identifiable subset of the larger parameter space, see Cheng and Traylor (1995). This lack of identifiability leads to the degeneracy of the Fisher information of the model, so that the classical χ^2 theory does not apply.

Of substantial interest is testing for homogeneity ($m = 1$) against heterogeneity ($m > 1$). For one-parameter families fulfilling some regularity conditions, e.g. the Poisson family, Chen, Chen and Kalbfleisch (2001) developed a modified likelihood ratio test (MLRT) and showed that the asymptotic distribution of this test statistic is a mixture of χ_0^2 and χ_1^2, with χ_0^2 being a point mass in 0 and χ_p^2 denoting the χ^2 distribution with $p > 0$ degrees of freedom. Recently, Chen and Li (2009) investigated the so called EM-test for testing for homogeneity in normal mixture models with distinct means and distinct variances. They showed that under conditions and the hypothesis of one component, the asymptotic distribution of the EM-test statistic is a χ_2^2 distribution. This test has been used by Vollmer, Holzmann, Ketterer and Klasen (2010) to analyze the distribution of annual (log) GDP per employee in Germany after reunification and to show that there are still two components. Another way of testing for homogeneity in mixture models would be to test the hypothesis that the observations are i.i.d. against a not explicitly given alternative using goodness-of-fit tests, see e.g. Cheng and Traylor (1995). For normal mixture models the BHEP-test would be an appropriate choice. For an overview about tests for normality, see Henze (2002).

Testing the hypothesis

$$H : m = m_0 \text{ versus } K : m = m_0 + 1, \tag{1.1.2}$$

can be performed via a time-demanding Bootstrap-approach, which has been introduced by McLachlan (1987). For one-parameter mixture models, Chen and Li (2010) recently developed an EM-test for testing the hypothesis (1.1.2) and showed that under regularity conditions on the underlying parametric family, the asymptotic distribution of the test statistic is a mixture of $\chi^2_0, \ldots, \chi^2_{m_0}$ distributions where the weights of the mixture distribution can easily be computed.

Another way for choosing the number of components are (penalized) likelihood based methods such as AIC and BIC. Leroux (1992a) showed that under certain weak conditions, penalized log likelihood criteria such as the BIC do not underestimate the true number of components, asymptotically. Actually, Gassiat (2002) shows that the true number of components is not overestimated, asymptotically, as well.

1.2 Hidden Markov models

A *hidden Markov model* (HMM) is a bivariate process $(S_k, X_k)_k$, where $(S_k)_k$ is an unobservable Markov chain with values in a finite space $\mathcal{M} = \{1, \ldots, m\}$ and the observable process $(X_k)_k$ is a process with values in a measurable set $\mathcal{X}$. Conditional on $(S_k)_k$, $(X_k)_k$ is a sequence of independent random variables such that the conditional distribution of X_t depends on S_t only. HMMs extend finite mixture models to deal with time series data that exhibit dependence over time. They relax the assumption that the hidden variable S_t is an i.i.d. $Mult(1; \boldsymbol{\alpha})$ random variable. Instead, one models the hidden process by a Markov chain. The dependence structure of an HMM can be represented by a directed graph (see Figure 1.1). Applications of hidden Markov models are to be found in the field of speech processing, genetics or financial economics. We refer to Zucchini and MacDonald (2009) for a comprehensive treatment, including applications of HMMs, and references therein.

Often the unobservable (or hidden) process $(S_k)_k$ is called *regime*. Calling the realizations of finite Markov chains states, the conditional distribution functions $P(X_k \leq x | S_k = j)$, $j = 1, \ldots, m$, are *state-dependent distribution functions* (abbreviated sdfs). Usually, the sdfs come from a parametric familiy $(F_\vartheta)_{\vartheta \in \Theta}$, e.g. normal distribution (see e.g. Cappé, Moulines and Rydén, 2005). In this case, the parameter of interest of the model $\boldsymbol{\omega}$ consists of the entries of the transition probability matrix of the hidden Markov chain $P_{\boldsymbol{\omega}} = (a_{ij})_{1 \leq i,j \leq m}$ and the parameters $\vartheta_1, \ldots, \vartheta_m$ of the sdfs. Note that *finite mixture models* are a special case of HMMs, since independence in the Markov chain can be expressed via transition

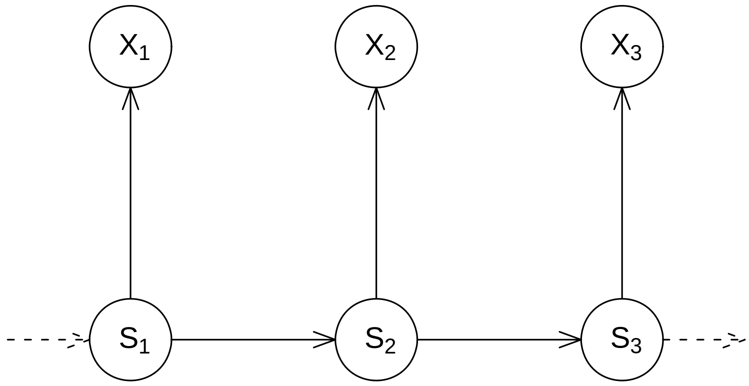

Figure 1.1: Dependency structure of a (basic) HMM. Here $(X_k)_k$ is the observable process and $(S_k)_k$ is the hidden Markov chain.

probabilities which do not depend on the initial state, i.e. $a_{1j} = \ldots = a_{mj}$ for all $j = 1, \ldots, m$.

Assuming that the hidden Markov chain $(S_k)_k$ is stationary and ergodic, so that the stationary distribution $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$ of the associated transition probability matrix $P_{\boldsymbol{\omega}}$ is uniquely determined, the marginal distribution of each X_k is given by the finite mixture

$$G(x) = \alpha_1 F_{\vartheta_1}(x) + \ldots + \alpha_m F_{\vartheta_m}(x).$$

Based on this marginal distribution, model selecting criteria and tests have been developed for HMMs by e.g. Poskitt and Zhang (2005) or Dannemann and Holzmann (2008).

Identifiability in HMMs

As for finite mixtures identifiability is an important issue in the HMM framework. For HMMs with sdfs from the same parametric family, Leroux (1992b) shows how an argument of Teicher (1967) can be used to establish identifiability if it is assumed to hold for the corresponding finite mixture. Therefore, HMMs with e.g. Gaussian, gamma or Poisson distributions as sdfs are identifiable. For a short overview about identifiability in HMMs, including the illustrating Example 12.4.5 of Gaussian HMMs, see Cappé et al. (2005). Since homoscedastic as well as normal mixtures with possibly distinct means and variances are identifiable they assume for notational simplicity that the sdfs are given by $P(X_k \leq x | S_k = i) = \Phi\big((x - \mu_i)/\sigma\big)$, i.e. σ is a structural parameter.

Parameter Estimation

Usually, the parameters of an HMM are estimated using maximum likelihood. Following Douc, Moulines and Rydén (2004) we consider the (log) likelihood conditional on $S_0 = i_0$,

$$\begin{aligned}\tilde{l}_n(\boldsymbol{\omega}) &= \tilde{l}_n(\boldsymbol{\omega}; X_1, \dots, X_n | S_0 = i_0) \\ &= \log\Big(\sum_{i_1=1}^{m} \cdots \sum_{i_n=1}^{m} \prod_{k=1}^{n} a_{i_{k-1},i_k} \prod_{k=1}^{n} f_{\vartheta_{i_k}}(X_k)\Big) \\ &= \log \boldsymbol{e}_{i_0}^T \left(\prod_{k=1}^{n} P_{\boldsymbol{\omega}} G_{\boldsymbol{\omega}}(X_k)\right) \mathbf{1},\end{aligned} \tag{1.2.1}$$

where $P_\omega = (a_{ij})_{1 \le i,j \le m}$ is the transition probability matrix of the hidden Markov chain $(S_k)_k$, $G_{\boldsymbol{\omega}}(X_k) = \operatorname{diag}\big(f_{\vartheta_i}(X_k)_{i=1,\dots,m}\big)$, $\boldsymbol{e}_{i_0}$ is the i_0th unit vector of length m and $\mathbf{1} = (1, \dots, 1)^T$. Some researchers, e.g. Zucchini and MacDonald (2009), work with a slightly different version of (1.2.1). They do not condition on the state $S_0 = i_0$ but start with the initial distribution $\boldsymbol{\delta}$ of S_1. An appropriate choice for $\boldsymbol{\delta}$ is to choose $\boldsymbol{\delta} = \boldsymbol{\alpha}$, the stationary distribution of the hidden Markov chain, provided the latter exists. Since the log-likelihood equation has a highly nonlinear structure, there is no analytic solution for the ML estimates. Equation (1.2.1) shows that the log likelihood can be expressed as a product of matrices and therefore it can be easily evaluated. It can be maximized over $\boldsymbol{\omega}$ using standard numerical optimization procedures, such as Newton-type algorithms or the Nelder-Mead simplex algorithm, or using EM algorithm.

Model selection in HMMs

Selecting the number of states of the underlying hidden Markov chain is a task of major importance. To this end, model selection criteria such as BIC or AIC, based on the full-model log-likelihood (e.g. Zucchini and MacDonald, 2009) are often used. Poskitt and Zhang (2005) reduce the problem of determining the number of regimes in a stationary HMM to selecting the number of components of the marginal mixture distribution.

Testing for homogeneity in HMMs is more or less just of theoretical interest since in case of just one regime the observations $(X_k)_k$ are i.i.d. Therefore, the tests based on (modified) likelihood ratio developed for testing for homogeneity in mixture models (see e.g. Chen and Li, 2009) can be applied to HMMs, cf. Dannemann (2009, Sec. 3.2.2). Since one neglects the dependence structure under the alternative the power properties of these tests could be influenced. As the LRT statistic for testing for homogeneity in HMMs already diverges to infinity (see Gassiat and Keribin, 2000), there is just little hope to develop an asymptotic distribution theory for the more general problem of testing $m = 2$ against $m \geq 3$ in HMMs via the LRT. In the case that the underlying parametric family $(F_\vartheta)_{\vartheta \in \Theta}$ depends on a univariate parameter $\vartheta \in \Theta \subset \mathbb{R}$ and fulfills some regularity conditions Dannemann and

Holzmann (2008) developed a method for testing the hypothesis $m = 2$ against $m \geq 3$ by extending the modified likelihood ratio test (MLRT) for testing the hypothesis $m = 2$ against $m \geq 3$ in mixture models based on the marginal mixture distribution of an HMM. They show that the asymptotic distribution of the MLRT for HMMs is the same as for the corresponding finite mixture models.

1.3 Markov-switching autoregressive models

A *Markov-switching autoregressive model* is a bivariate process $(S_k, X_k)_k$, where $(S_k)_k$ is a Markov chain with values in a finite space $\mathcal{M} = \{1, \ldots, m\}$ and, conditional on $(S_k)_k$, $(X_k)_k$ is an inhomogeneous p-order Markov chain on a state space $\mathcal{X}$ such that the conditional distribution of X_t only depends on S_t and lagged X's, say $X_{t-1}, \ldots, X_{t-p+1}$. The process $(S_k)_k$, usually referred to as the regime, is not observable and inference has to be carried out in terms of the observable process $(X_k)_k$. Here, we note that we will omit the prefix *Markov* when it is clear that we are dealing with *Markov-switching autoregressive models*. In this section we do not care about determining the number of regimes in the hidden Markov chain. We defer this discussion to the following chapters, especially to the beginning of Chapter 2.

In this thesis, we are concerned with two different classes of models:

(i) the *linear switching autoregressive* models, which are given in their most general form (see e.g. Sclove, 1983) by

$$X_t = \zeta_{S_t} + \phi_{1,S_t} X_{t-1} + \ldots + \phi_{p,S_t} X_{t-p} + \sigma_{S_t} \epsilon_t, \tag{1.3.1}$$

where $\epsilon_t \overset{iid}{\sim} D$ with $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = 1$.

(ii) the *switching ARCH* models which are given in their most general form (see e.g. Gray, 1996) by

$$X_t = \sigma_t \epsilon_t; \quad \sigma_t^2 = \vartheta_{S_t} + \phi_{1,S_t} X_{t-1}^2 + \ldots + \phi_{p,S_t} X_{t-p}^2 \tag{1.3.2}$$

where $\epsilon_t \overset{iid}{\sim} D$ with $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = 1$.

Here we assume that $D = N(0,1)$ or $D = t(\nu)$ and denote by $t(\nu)$ the (standardized) t-distribution with $\nu > 2$ degrees of freedom and variance 1.

In general we write such a model as

$$X_t = F_{\boldsymbol{\omega}}(X_{t-1}^p, S_t; \epsilon_t), \tag{1.3.3}$$

where $(F_{\boldsymbol{\omega}})_{\boldsymbol{\omega}}$ is a family indexed by a finite dimensional parameter $\boldsymbol{\omega}$, $(\epsilon_k)_k$ is an independent and identically distributed sequence of random variables with $E(\epsilon_1) = 0$ and $E(\epsilon_1^2) = 1$ and $X_k^p = (X_k, \ldots, X_{k-p+1})$. In Section 1.5 we specify this model for a two-state Markov chain and discuss the entries contained in $\boldsymbol{\omega}$.

The dependence structure of a *Markov-switching autoregressive model* can be represented by a directed graph. Figure 1.2 states this dependence graph for a *Markov-switching autoregressive model* of order 1. The nodes (circles) correspond to the random variables and the edges (arrows) represent the structure of the joint probability distribution. Thus, Figure 1.2 implies that the distribution of a random variable S_t conditional on the history of the process $S_{t-1}, S_{t-2}, \dots$ is completely determined by the value of its predecessor S_{t-1}. This is exactly the property that $(S_k)_k$ forms a (first order) Markov chain. The distribution of X_t conditional on the past observations $X_{t-1}, X_{t-2}, \dots$ and the states $S_t, S_{t-1}, \dots$ is determined by S_t and X_{t-1} and this is exactly the property we postulate on a Markov-switching autoregressive model (with $p = 1$). Here, we note that the model which was

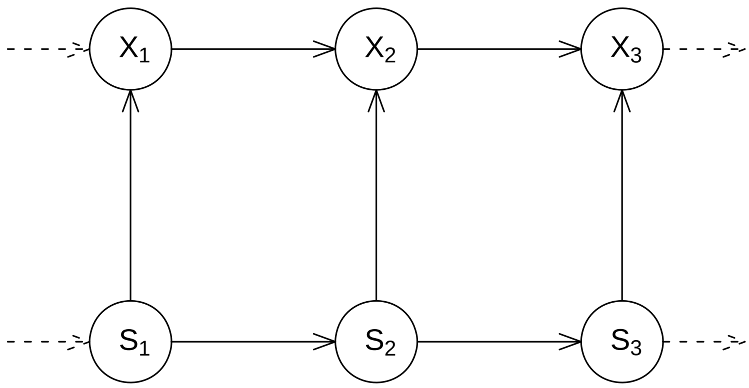

Figure 1.2: Dependency structure of a Markov-switching autoregressive model, where $(X_k)_k$ is the observable process and $(S_k)_k$ is the hidden Markov chain.

introduced by Hamilton (1989), where the mean level in the linear switching autoregressive model switches,

$$X_t - \mu_{S_t} = \phi_1(X_{t-1} - \mu_{S_{t-1}}) + \ldots + \phi_p(X_{t-p} - \mu_{S_{t-p}}) + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0, \sigma^2), \qquad (1.3.4)$$

does not match our model specification (1.3.3) since the distribution of X_t does not depend on S_t only but also on $S_{t-1}, \dots, S_{t-p}$. One possibility to overcome this problem is to introduce a multivariate state vector $\boldsymbol{S}_t = (S_t, \dots, S_{t-p})$ as in Frühwirth-Schnatter (2006). Then $(\boldsymbol{S}_k)_k$ is a first order Markov chain on $\mathcal{M}^{p+1}$. Testing for the number of states of $(S_k)_k$ or $(\boldsymbol{S}_k)_k$ would be sophisticated, though.

As mentioned in Cappé et al. (2005), it is not clear if there exists a strictly stationary solution of equation (1.3.3) for any given parameter $\boldsymbol{\omega}$ and innovations $(\epsilon_k)_k$. For the models (1.3.1) and (1.3.2) we give sufficient conditions for the existence of such solutions which are due to Francq and Zakoïan (2001) and Stelzer (2005, 2009).

Linear switching autoregression

Assuming that $(S_k)_k$ is an irreducible, aperiodic Markov chain starting from its ergodic distribution $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)$, we give some sufficient conditions for the existence of a strictly stationary solution $(X_k)_k$ of (1.3.1). In order to investigate the properties of strict stationarity we write (1.3.1) as a stochastic recurrence equation of the form

$$\mathbf{X}_t = \mathbf{A}_t \mathbf{X}_{t-1} + \mathbf{c}_t \tag{1.3.5}$$

with $\mathbf{X}_t := (X_t, \dots, X_{t-p+1})^T \in \mathbb{R}^p$, $p \geq 1$, $\mathbf{c}_t := (\zeta_{S_t} + \sigma_{S_t}\epsilon_t, 0, \dots, 0)^T \in \mathbb{R}^p$ and

$$\mathbf{A}_t := \begin{pmatrix} \phi_{1,S_t} & \phi_{2,S_t} & \cdots & \cdots & \phi_{p,S_t} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \ddots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \in \mathbb{R}^{p \times p}.$$

Let $\|\cdot\|$ denote any norm on $\mathbb{R}^p$, write

$$\|\mathbf{A}\| = \max_{\mathbf{x} \in \mathbb{R}^p \setminus \{\mathbf{0}\}} \frac{\|\mathbf{A}\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$$

for the induced matrix norm, and put $\log^+(x) = \max\{\log(x), 0\}$, $x > 0$. It is clear that $E \log^+ \|\mathbf{A}_t\| < \infty$ and $E \log^+ \|\mathbf{c}_t\| < \infty$, since the state space of the hidden Markov chain $(S_k)_k$ is finite. From Brandt (1986), the unique stationary solution of (1.3.5) is given by

$$\mathbf{X}_t = \mathbf{c}_t + \sum_{k=1}^{\infty} \mathbf{A}_t \mathbf{A}_{t-1} \dots \mathbf{A}_{t-k+1} \mathbf{c}_{t-k},$$

whenever the top Lyapunov exponent γ, defined by

$$\gamma := \inf_{t \geq 1} E \frac{1}{t} \log \|\mathbf{A}_t \mathbf{A}_{t-1} \dots \mathbf{A}_1\|,$$

is strictly negative. Obviously, any strictly stationary solution $(X_k)_k$ of (1.3.1) leads to a strictly stationary solution of (1.3.5) via the above transformation. On the other hand we can see that the first component of the strictly stationary solution of (1.3.5) leads to a strictly stationary solution of (1.3.1).

In the case of purely deterministic AR-coefficients, such as in models (2.2.1) or (2.2.3) we can give an equivalent condition: Denoting by $\rho(\mathbf{A}_0)$ the spectral radius of $\mathbf{A}_0$, then $\gamma < 0$ if and only if $\rho(\mathbf{A}_0) = \lim_{n\to\infty} \sqrt[n]{\|\mathbf{A}_0^n\|} < 1$ which in turn holds if and only if $\det(\boldsymbol{I}_p - z\mathbf{A}_0) = 1 - \phi_1 z - \dots \phi_p z^p \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$. The latter is just the sufficient condition for classical ARMA processes to be causal (see e.g. Brockwell and

Davis, 2006, Thm. 3.1.1).

In the case of a linear switching autoregressive model with switching intercept, see e.g. model (2.2.1), Krolzig (1997) gives an ARMA representation for this model. Note that a weakly stationary process is said to admit an ARMA(p,q) representation if it has the same autocovariance structure as a causal and invertible ARMA(p,q) process, i.e. if and only if its autocovariances satisfy a difference equation of minimal order p with minimal rank $q+1$, see e.g. Zhang and Stine (2001).

Switching ARCH

Assuming that $(S_k)_k$ is an irreducible, aperiodic Markov chain starting from its ergodic distribution $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$, we give some sufficient conditions for the existence of a strictly stationary solution $(X_k, \sigma_k^2)_k$ of (1.3.2). In order to investigate the properties of strict stationarity we write the squared form of the Markov switching ARCH equations (1.3.2)

$$X_t^2 = \sigma_t^2 \epsilon_t^2; \quad \sigma_t^2 = \vartheta_{S_t} + \phi_{1,S_t} X_{t-1}^2 + \ldots + \phi_{p,S_t} X_{t-p}^2 \tag{1.3.6}$$

in the form of a stochastic recurrence equation

$$\mathbf{X}_t = \mathbf{A}_t \mathbf{X}_{t-1} + \mathbf{c}_t. \tag{1.3.7}$$

Without loss of generality let $p \geq 2$. This assumption is not very restrictive since we can always use the representation below by simply including higher order terms with ARCH coefficients equal to zero. Let $\mathbf{X}_t := (\sigma_{t+1}^2, \sigma_t^2, X_t^2, \ldots, X_{t-p+2}^2)^T \in (\mathbb{R}_{\geq 0})^{p+1}$, $\mathbf{c}_t := (\vartheta_{S_{t+1}}, 0, \ldots, 0)^T \in (\mathbb{R}_{\geq 0})^{p+1}$ and

$$\mathbf{A}_t := \begin{pmatrix} \phi_{1,S_{t+1}}\epsilon_t^2 & 0 & \phi_{2,S_{t+1}} & \cdots & \cdots & \phi_{p,S_{t+1}} \\ 1 & 0 & 0 & \cdots & \cdots & 0 \\ \epsilon_t^2 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \in \mathbb{R}^{p+1,p+1}.$$

Since the state space of the hidden Markov chain $(S_k)_k$ is finite, $E \log^+ \|\mathbf{A}_t\| < \infty$ and $E \log^+ \|\mathbf{c}_t\| < \infty$. From Brandt (1986), the unique stationary solution of (1.3.7) is given by

$$\mathbf{X}_t = \mathbf{c}_t + \sum_{k=1}^{\infty} \mathbf{A}_t \mathbf{A}_{t-1} \ldots \mathbf{A}_{t-k+1} \mathbf{c}_{t-k},$$

whenever the top Lyapunov exponent γ is strictly negative. Obviously, any strictly stationary solution $(X_k^2, \sigma_k^2)_k$ of (1.3.6) leads to a strictly stationary solution of (1.3.5) via the above transformation. On the other hand, we can see that the second and the third component of the strictly stationary solution of (1.3.7) lead to a strictly stationary solution of (1.3.6). The unique strictly stationary solution $(X_k, \sigma_k^2)_k$ of (1.3.2) is formed by $X_k = \sqrt{\sigma_k^2}\epsilon_k$ and the second coordinate of the strictly stationary solution of (1.3.7), see Stelzer (2005, Thm. 6.3).

In case of purely deterministic autoregressive parameters in the ARCH-part of (1.3.2) as in model (2.2.4), the sufficient condition that (1.3.2) admits a stationary solution is the same as for non-switching ARCH processes (see e.g. Bougerol and Picard, 1992).

1.4 Related models

Self-exciting threshold autoregressive models

Self-exciting threshold autoregressive models (SETAR), introduced by Tong (1983), are closely related to *(linear) Markov-switching autoregressive models*. Both models are designed to capture discrete changes in the series that generate the data. While in *Markov-switching autoregressive models* the movement between regimes is unrelated to the past observations of the process and the regime is an unobservable process, movement between regimes in the SETAR model depends on the past observations of the process $(X_k)_k$. Regime switches occur according to the level of a theshold variable $Z_t = X_{t-\bar{d}}$, where $\bar{d} > 0$ is the so called delay parameter. In the following we state a SETAR model with 2 states (with an obvious extension to more than 2 states)

$$X_t = \begin{cases} \zeta_1 + \phi_{1,1}X_{t-1} + \ldots + \phi_{p,1}X_{t-p} + \sigma_1\epsilon_t, & \text{if } X_{t-\bar{d}} \le \tau, \\ \zeta_2 + \phi_{1,2}X_{t-1} + \ldots + \phi_{p,2}X_{t-p} + \sigma_2\epsilon_t, & \text{if } X_{t-\bar{d}} > \tau \end{cases} \tag{1.4.1}$$

with $\epsilon_t \overset{iid}{\sim} D$, e.g. $D = N(0,1)$. Since the delay parameter $\bar{d}$ and the threshold parameter τ are not observable one has to estimate them together with the other parameters. For a short review about parameter estimation in SETAR models, including Bayesian approaches, see Potter (1999). For a short overview about SETAR models and their relations to *Markov-switching autoregressive models*, see the monographs in Frühwirth-Schnatter (2006, Chp. 12.2) and Piger (2009).

Mixture autoregressive models

A popular subclass of *Markov-switching autoregressive models* are the so called *mixture autoregressive models* which were introduced in Juang and Rabiner (1985). In two articles Wong and Li (2000, 2001) considered linear mixture autoregressive models as well as mixture AR-ARCH models. These models result as a special case of *Markov-switching au-*

toregressive models where for the hidden process $S_t \overset{iid}{\sim} Mult(1; \boldsymbol{\alpha})$ holds. One advantage of

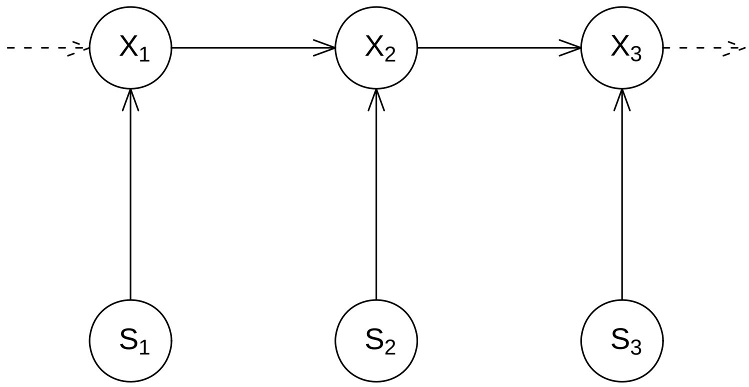

Figure 1.3: Dependency structure of a mixture autoregressive model of order 1, where $(X_k)_k$ is the observable process and $(S_k)_k$ are the hidden random variables.

this model is that the one-step-ahead predictor can be computed easily. Beyond that, Boshnakov (2006) showed that the multistep-predictors in linear mixture autoregressive models are also mixture distributions when the innovations are normal or more general α-stable distributed by deriving the conditional characteristic function $\varphi_{t+h|t}(s) = E\big(\exp(isX_{t+h})|\mathcal{F}_t\big)$, where $\mathcal{F}_t$ is the σ-algebra generated by $X_t, X_{t-1}, \ldots$ Even though *mixture autoregressive models* have some nice properties one also has to mention that there is a crucial drawback of these models: The autocorrelation in X_t is introduced only through the observation equation. Therefore these models are not able to capture spurious autocorrelation that disappears when conditioning on the state S_t.

1.5 Standing assumptions and methodology

The latent process

We assume that $(S_k)_k$ is a stochastic process with values in $\mathcal{M} = \{1, \ldots, m\}$. Throughout the thesis we assume that the Markov chain is time homogeneous, i.e. the transition probabilities $a_{ij} = P(S_k = j|S_{k-1} = i)$, $1 \leq i, j \leq m$, do not depend on k. Moreover, we assume that the Markov chain $(S_k)_k$ is irreducible and aperiodic. This condition ensures that $(S_k)_k$ is an ergodic process with unique stationary distribution $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$ with $\alpha_k > 0$, $k = 1, \ldots, m$. We usually assume that the initial distribution $\boldsymbol{\delta}$ coincides with $\boldsymbol{\alpha}$. Together with the homogeneity assumption this assures the stationarity of the process $(S_k)_k$.

Testing for homogeneity in Markov-swichting autoregressive models

Testing for regime switching in Markov-switching autoregressive models corresponds to testing the null hypothesis $\mathcal{M} = \{1\}$ of a single state (so that the model reduces to a mere autoregressive process) against the alternative hypothesis $\mathcal{M} = \{1, 2\}$ of (at least) two states. Deriving the asymptotic distribution of the LRT and related test statistics is a difficult task for a variety of reasons. First, under the null hypothesis, parameters of the full model are not identifiable, and the asymptotic distribution of the corresponding LRT will be highly non-standard. This problem already arises in the closely related problem of testing for homogeneity in two-component mixtures, which has been intensively studied in recent years, see Chen et al. (2001), Dacunha-Castelle and Gassiat (1999) and Liu and Shao (2003), see also Andrews (2001). Second, additional difficulties arise if the Markov dependence structure of the regime is incorporated into the test statistic. Indeed, even for compact parameter spaces, Gassiat and Kerebin (2000) show that the LRT for regime switching may not converge in distribution at all.

Therefore, Cho and White (2007) suggest a quasi LRT for switching regime in general autoregressive models which neglects the dependence structure of the regime under the alternative, and derive its asymptotic distribution. When testing for the presence of k against more than k states ($k \geq 2$), this approach generally also affects the asymptotic distribution under the hypothesis (however cf. Dannemann and Holzmann 2008), but when testing for the presence of regime switching, there is no regime under the hypothesis, and hence this approach only affects the power properties.

We extend the approach by Cho and White (2007) along the lines of Chen et al. (2001) and obtain a penalized (or modified) quasi LRT with an easily tractable asymptotic distribution and comparable power properties to the quasi LRT.

As noted in Section 1.3, we write a switching autoregressive model of the form

$$X_t = F_{\boldsymbol{\omega}}(S_t, X_{t-1}^p; \epsilon_t), \tag{1.5.1}$$

with innovations $(\epsilon_k)_k$ and $(F_{\boldsymbol{\omega}})_{\boldsymbol{\omega}}$ being a family of functions indexed by some finite-dimensional parameter $\boldsymbol{\omega}$. For a two-state chain $(S_k)_k$, $\boldsymbol{\omega}$ consists of the entries a_{21}, a_{12} of the transition matrix $P_{\boldsymbol{\omega}} = (a_{ij})_{i,j=1,2}$, the switching parameters $\vartheta_1, \vartheta_2 \in \Theta \subset \mathbb{R}^r$ as well as the structural parameters $\boldsymbol{\eta} \in \boldsymbol{H} \subset \mathbb{R}^d$ which are the same for all states, so that

$$\boldsymbol{\omega}^T = \big(a_{21}, a_{12}, \vartheta_1, \vartheta_2, \boldsymbol{\eta}^T\big).$$

In Chapter 2 and 3, we assume $r = 1$ whereas we assume the switching parameter to be bivariate in Chapter 4.

Penalized maximum likelihood estimation

Likelihood based methods play a prominent role for parameter estimation in switching autoregressive models. Suppose that conditional on $X_{k-1}^p = x_{k-1}^p$ and $S_k = i$, X_k has density $g(x_k|x_{k-1}^p; \vartheta_i, \boldsymbol{\eta})$ w.r.t. some σ-finite measure μ on $\mathcal{X}$. Then the conditional likelihood given the initial observations $X_0^p = (X_0, \ldots, X_{-p+1})$ (we start indexing from $-p+1, -p+2, \ldots$) and the initial unobserved state $S_0 = i_0$ is given by

$$\begin{aligned}\tilde{l}_n(\boldsymbol{\omega}) &= \log\Big(\sum_{i_1=1}^{2}\cdots\sum_{i_n=1}^{2}\prod_{k=1}^{n} a_{i_{k-1},i_k}\prod_{k=1}^{n} g(X_k|X_{k-1}^p; \vartheta_{i_k}, \boldsymbol{\eta})\Big)\\ &= \log \boldsymbol{e}_{i_0}^T\left(\prod_{k=1}^{n} P_{\boldsymbol{\omega}} G_{\boldsymbol{\omega}}(X_k^{p+1})\right)\mathbf{1},\end{aligned} \tag{1.5.2}$$

where $P_\omega = (a_{ij})_{1\le i,j\le 2}$ is the transition probability matrix of the hidden Markov chain $(S_k)_k$, $G_{\boldsymbol{\omega}}(X_k^{p+1}) = \operatorname{diag}\big(g(X_k|X_{k-1}^p; \vartheta_i, \boldsymbol{\eta})_{i=1,2}\big)$, $\boldsymbol{e}_{i_0}$ is the i_0th unit vector of length 2 and $\mathbf{1} = (1,1)^T$. Here we note, that we can condition on $S_0 = i_0$ without loss of generality, since for any model with initial state $S_0 = i_0' \neq i_0$, we can find an equivalent model with initial state i_0 by relabeling the states of the hidden Markov chain and reordering the a_{ij}'s and ϑ_i's accordingly.

The maximizer $\hat{\boldsymbol{\omega}}$ of $\tilde{l}_n(\boldsymbol{\omega})$ is called the (conditional) maximum likelihood estimate. Its asymptotic properties, especially consistency as well as asymptotic normality are well-established by now (cf. Douc et al. 2004).

As indicated above, instead of using the full-model log likelihood function $\tilde{l}_n(\boldsymbol{\omega})$ we shall base inference on a quasi likelihood which neglects the dependence structure in the regime. Let $\boldsymbol{\psi}^T = (\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}^T)$,

$$g_{\text{mix}}(x_t|x_{t-1}^p; \boldsymbol{\psi}) = (1-\alpha) g(x_t|x_{t-1}^p; \vartheta_1, \boldsymbol{\eta}) + \alpha g(x_t|x_{t-1}^p; \vartheta_2, \boldsymbol{\eta}) \tag{1.5.3}$$

and consider the *quasi log-likelihood function* given by

$$l_n(\boldsymbol{\psi}) = \sum_{t=1}^{n} \log g_{\text{mix}}(X_t|X_{t-1}^p; \boldsymbol{\psi}). \tag{1.5.4}$$

Remark 1.1. Note that (1.5.4) is the true likelihood function only if the regime is independent. For the time series model itself, an independent regime may not appear particularly attractive (as seen in Section 1.4), but it can nevertheless be used for constructing a feasible test for regime switching. For a Markov-dependent regime, the parameter $(1-\alpha, \alpha)$ in (1.5.3) corresponds to the stationary distribution of the underlying transition matrix.

Following Chen et al. (2001, 2004) and Chen and Li (2009), in order to obtain a feasible asymptotic distribution we consider a penalized version of l_n, called *modified* or *penalized* *quasi likelihood function*, which is defined by

$$pl_n(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) = l_n(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) + p(\alpha), \tag{1.5.5}$$

where $p(\alpha)$ is a penalty with the following properties:

(i) $p(\alpha)$ attains its maximum at $\alpha = 0.5$,

(ii) $p(\alpha)$ is continuous on $(0, 1)$,

(iii) $p(\alpha) = p(1 - \alpha)$ and

(iv) $p(\alpha) \to -\infty$ for $\alpha \to 0$.

Examples are

$$p(\alpha) = C \, \log\big(4\,\alpha(1-\alpha)\big) \qquad \text{or} \qquad p(\alpha) = C \, \log(1 - |1 - 2\alpha|). \tag{1.5.6}$$

In the following we develop tests for homogeneity in Markov-switching autoregressive models based on the penalized quasi likelihood function (1.5.5).

2 Feasible Tests for regime switching in autoregressive models

Estimating the true number of regimes in a switching autoregressive model is a task of major importance, see e.g. Olteanu and Rynkiewicz (2007). One attempt for testing the hypothesis

$$H : m = m_0 \text{ versus } K : m = m_0 + 1,$$

where m denotes the number of regimes in the hidden Markov chain is the so called Bootstrap-approach which has been introduced by McLachlan (1987) in the case of mixture models. As is well known this approach is computationally very intensive since it requires repeated maximization of the full likelihood under the hypothesis and the alternative. Furthermore, we should keep in mind that the asymptotic correctness of such bootstrap tests has not been established yet and is far from being obvious.

For linear mixture autoregressive models Naik, Shi and Tsai (2007) introduced a new information criterion (*mixture regression criterion*, abbreviated MRC) for jointly determining the order of the autoregressive process p and the number of components m. An extension of their criterion to Markov-switching autoregressive models seems to be possible, see Dupont (2010).

In this chapter we develop methods for testing for homogeneity in several switching autoregressive models.

2.1 Testing for the number of components in a Markov-switching autoregressive model

In this chapter we are concerned with the basic methodological issue to determine the number of states of the underlying regime in a switching autoregressive model, or in a first place to test for the existence of at least two states. Major progress on the topic was recently made by Cho and White (2007), who derive the asymptotic distribution of a quasi likelihood ratio test (quasi LRT) (which neglects the serial dependence of the regime under the alternative). The resulting asymptotic distribution is quite involved, however, and depends both on the underlying parametric model as well as on the true parameter values. Therefore, following Chen, Chen and Kalbfleisch (2001) and Chen and Li (2009) for i.i.d. mixtures, we propose a penalized version of the test statistics and obtain a simple asymptotic distribution, a mixture of a point mass at zero and a χ_1^2-distribution with equal weights. Further, simulations indicate that this does not result in any loss of power in finite samples as compared to the original quasi LRT.

Since the seminal paper by Hamilton (1989), who introduced regime switching autoregressive models and used them for business cycle analysis of U.S. GNP data, these models were applied to a variety of economic data including macroeconomic time series (e.g. Porter 1983 for investigating cartel behavior; Davig 2004 for the U.S. debt-output ratio) and financial time series (Hamilton and Susmel 1994 for stock returns or Cai 1994 for treasury bills) and are also frequently used in other areas such as electrical engineering. See Hamilton (2008) for a recent survey article.

The outline of this chapter is as follows. In Section 2.2 we specify the model, give some examples and discuss consistency properties of penalized quasi maximum likelihood estimators. Section 2.3 deals with the asymptotic distributions of the penalized (or modified) quasi LRT and a related test called the EM-test (cf. Chen and Li 2009). In Section 2.4 we report the results of a simulation study. Proofs are deferred to Section 2.5.

2.2 Examples and estimation

2.2.1 Markov-switching autoregressive models

In Chapter 1 we introduced *Markov-switching autoregressive models*. Assuming that $(S_k)_k$ takes values in $\mathcal{M} = \{1, 2\}$, we write a switching autoregressive model of the form

$$X_t = F_{\boldsymbol{\omega}}(S_t, X_{t-1}^p; \epsilon_t),$$

with innovations $(\epsilon_k)_k$ and $(F_{\boldsymbol{\omega}})_{\boldsymbol{\omega}}$ being a family of functions indexed by some finite-dimensional parameter

$$\boldsymbol{\omega}^T = \left(a_{21}, a_{12}, \vartheta_1, \vartheta_2, \boldsymbol{\eta}^T\right)$$

with switching parameters $\vartheta_1, \vartheta_2 \in \Theta \subset \mathbb{R}^r$ as well as the structural parameters $\boldsymbol{\eta} \in \boldsymbol{H} \subset \mathbb{R}^d$. The parameter sets Θ and $\boldsymbol{H}$ are assumed to be compact. In this chapter we discuss testing for homogeneity in various switching autoregressive models with univariate switching parameter, i.e. $r = 1$.

Example 2.1 (*Linear switching autoregression*)**.** 1. The linear switching autoregressive model with switching intercept is given by

$$X_t = \zeta_{S_t} + \sum_{j=1}^{p} \phi_j X_{t-j} + \sigma \epsilon_t, \tag{2.2.1}$$

where σ is a scale parameter for the innovation distribution, the ϕ_j's are the (non-switching) autoregressive parameters, and the intercept ζ switches according to S_t. Krolzig (1997) and Hamilton (2008) give further motivation and discussion of the properties. Compared to model (1.3.1) in which all parameters are affected by the hidden state S_t, we confine ourselves in this model to an intercept which is state-dependent, whereas the autoregressive parameters as well as the scale parameter of the innovations is equal for every regime, i.e. these parameters are structural parameters. This model allows for shifts in the mean level and assumes that the dynamic pattern of the time series is equal for both states. For the innovations, the normal distribution is a standard choice (cf. Cho and White 2007); another useful distribution is the t-distribution, which allows for thicker tails which are often observed empirically.

In the above notation, we have $\vartheta_i = \zeta_i$, $i = 1, 2$. If σ is fixed, we have $d = p$ and $\boldsymbol{\eta} = (\phi_1, \dots, \phi_p)^T$, otherwise, $d = p + 1$ and $\boldsymbol{\eta} = (\phi_1, \dots, \phi_p, \sigma)^T$.

2. The linear switching autoregressive model with one switching autoregressive parameter is given by

$$X_t = \zeta + \sum_{j=1}^{j_0-1} \phi_j X_{t-j} + \phi_{j_0,S_t} X_{t-j_0} + \sum_{j_0+1}^{p} \phi_j X_{t-j} + \sigma\epsilon_t, \tag{2.2.2}$$

where ζ is the non-switching intercept, ϕ_j, $j = 1, \dots, j_0 - 1, j_0 + 1, \dots, p$, are the (non-switching) autoregressive parameters, σ is the scale parameter of the innovation process and ϕ_{j_0,S_t} switches according to S_t. In contrast to model (1.3.1), this model does not allow for different mean levels or different scale parameters σ. It allows one autoregressive parameter to switch. Model (2.2.2) includes

$$X_t = \phi_{S_t} X_{t-1} + \sigma\epsilon_t,$$

which has been discussed in Lange and Rahbek (2009). Here, we consider t- as well as normal distributed innovations.

In the above notation, we have $d = p + 1$, $\boldsymbol{\eta} = (\zeta, \phi_1, \dots, \phi_{j_0-1}, \phi_{j_0+1}, \dots, \phi_p, \sigma)^T$ and $\vartheta_i = \phi_{j_0,i}$, $i = 1, 2$.

3. The linear switching autoregressive model with switching variance is given by

$$X_t = \zeta + \sum_{j=1}^{p} \phi_j X_{t-j} + \sigma_{S_t}\epsilon_t, \tag{2.2.3}$$

where σ is a scale parameter for the innovation distribution which switches according to S_t, the intercept ζ and the ϕ_j's are the non-switching parameters. This model is very popular for time series of asset prices (see e.g. Piger 2009). In this model only the scale parameter σ is affected by the hidden state S_t, whereas all the parameters are allowed to switch in model (1.3.1). This model captures switches in the volatility and is not able to capture shifts in the mean level. Again we consider t-distributed as well as normal distributed innovations.

In the above notation, we have $d = p + 1$, $\vartheta_i = \sigma_i$, $i = 1, 2$, and $\boldsymbol{\eta} = (\zeta, \phi_1, \dots, \phi_p)^T$.

Example 2.2 (*Switching ARCH)***.** Regime switching ARCH-models were introduced by Hamilton and Susmel (1994) and by Cai (1994). The model specification by Cai (1994) when neglecting leverage effects is (Hamilton and Susmel 1994, give a slightly different specification)

$$X_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \vartheta_{S_t} + \sum_{j=1}^{p} \phi_j X_{t-j}^2, \tag{2.2.4}$$

with parameters $\vartheta_i \geq 0$, $i = 1, 2$, and $\phi_j \geq 0$, $j = 1, \ldots, p$. Compared to model (1.3.2), model (2.2.4) allows only the intercept to switch according to S_t. Nevertheless, it is able to take into account sudden changes in the volatility.

In the above notation, we have $d = p$ and $\boldsymbol{\eta} = (\phi_1, \ldots, \phi_p)^T$. Again, we consider both normal as well as t-distributed (cf. Tsay 2002) innovations.

2.2.2 Penalized maximum likelihood estimation

As indicated in Section 1.5, we base inference on a quasi likelihood which neglects the dependence structure in the regime and consider the *quasi log-likelihood function* which is given by

$$l_n(\boldsymbol{\psi}) = \sum_{t=1}^{n} \log g_{\mathrm{mix}}(X_t | X_{t-1}^p; \boldsymbol{\psi}),$$

where $\boldsymbol{\psi}^T = (\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}^T)$.

Following Chen et al. (2001, 2004) and Chen and Li (2009), in order to obtain a feasible asymptotic distribution we consider a penalized version of l_n, called *modified* or *penalized quasi likelihood function*, which is defined by

$$pl_n(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) = l_n(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) + p(\alpha), \tag{2.2.5}$$

where $p(\alpha)$ is a penalty function on α. Examples and properties of this penalty function are discussed in Section 1.5. Let $(\widehat{\alpha}, \widehat{\vartheta}_1, \widehat{\vartheta}_2, \widehat{\boldsymbol{\eta}})$ (resp. $(\widehat{\alpha}^*, \widehat{\vartheta}_1^*, \widehat{\vartheta}_2^*, \widehat{\boldsymbol{\eta}}^*)$) be the maximizers of $l_n(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta})$ (resp. $pl_n(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta})$) over the parameter space $[0,1] \times \Theta^2 \times \boldsymbol{H}$, and let $(\widehat{\vartheta}_0, \widehat{\boldsymbol{\eta}}_0)$ be the maximizers of $l_n(1/2, \vartheta, \vartheta, \boldsymbol{\eta})$ or equivalently of $pl_n(1/2, \vartheta, \vartheta, \boldsymbol{\eta})$ over the parameter space $\Theta \times \boldsymbol{H}$. We denote the true parameter under the null hypothesis of no switching regime by $(\vartheta_0, \boldsymbol{\eta}_0)$. If not otherwise specified, we compute the probabilities and expectations with respect to this distribution. We shall need the following assumptions.

Assumption 2.1. The process $(\boldsymbol{Z}_k)_{k \geq 0} = (S_k, X_k, \ldots, X_{k-p+1})_{k \geq 0}$ is a Markov chain on $\mathcal{M} \times \mathcal{X}^p$. Under the null hypothesis, the observable process $(X_k)_k$ is strictly stationary and geometrically ergodic.

Assumption 2.2. *(Identifiability)* If for parameters $\boldsymbol{\psi}^T = (\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}^T)$ and $\boldsymbol{\psi}'^T =$

$(\alpha', \vartheta_1', \vartheta_2', \boldsymbol{\eta}'^T)$, $\alpha \notin \{0,1\}$ and $\vartheta_1 \neq \vartheta_2$, one has that

$$g_{\text{mix}}(x|\, y^p; \boldsymbol{\psi}) = g_{\text{mix}}(x|\, y^p; \boldsymbol{\psi}') \text{ for all } x \in \mathcal{X}, y^p \in \mathcal{X}^p,$$

then $\boldsymbol{\eta} = \boldsymbol{\eta}'$ and after possibly permuting the states of the Markov chain $(S_k)_k$, we further have that $\alpha = \alpha'$ and $\vartheta_i = \vartheta_i'$, $i = 1, 2$.

Assumption 2.3. For all fixed $x \in \mathcal{X}, y^p \in \mathcal{X}^p$, $g(x|y^p; \cdot, \cdot) \in C^{(2)}((\Theta, \boldsymbol{H}))$. Further, there exists a nonnegative function K such that

$$EK(X_1^{p+1}) < \infty \qquad \text{and} \quad |\log(g(x_1|x_0^p; \vartheta, \boldsymbol{\eta}))| \leq K(x_1^{p+1})$$

for all $x_1^{p+1} \in \mathcal{X}^{p+1}$ and all $(\vartheta, \boldsymbol{\eta}) \in \Theta \times \boldsymbol{H}$.

Define

$$\begin{aligned}
U_i^{\eta_j}(\boldsymbol{\eta}) &= \frac{g(X_i|X_{i-1}^p; \vartheta_0, \boldsymbol{\eta}) - g(X_i|X_{i-1}^p; \vartheta_0, \eta_1, \ldots, \eta_{j-1}, \eta_{j,0}, \eta_{j+1}, \ldots, \eta_d)}{\big(\eta_j - \eta_{j,0}\big) g(X_i|X_{i-1}^p; \vartheta_0, \boldsymbol{\eta}_0)}, \\
Y_i(\vartheta, \boldsymbol{\eta}) &= \frac{g(X_i|X_{i-1}^p; \vartheta, \boldsymbol{\eta}) - g(X_i|X_{i-1}^p; \vartheta_0, \boldsymbol{\eta})}{\big(\vartheta - \vartheta_0\big) g(X_i|X_{i-1}^p; \vartheta_0, \boldsymbol{\eta}_0)}, \quad Y_i(\vartheta) = Y_i(\vartheta, \boldsymbol{\eta}_0), \\
Z_i(\vartheta) &= \frac{Y_i(\vartheta, \boldsymbol{\eta}_0) - Y_i(\vartheta_0, \boldsymbol{\eta}_0)}{\vartheta - \vartheta_0}, \quad Z_i = \frac{\partial}{\partial \vartheta} Y_i(\vartheta)\bigg|_{\vartheta = \vartheta_0}
\end{aligned}$$

and

$$\begin{aligned}
U_i^{\eta_j}(\eta_1, \ldots, \eta_{j-1}, \eta_{j,0}, \eta_{j+1}, \ldots, \eta_d) &= \frac{\frac{\partial}{\partial \eta_j} g(X_i|X_{i-1}^p; \vartheta_0, \boldsymbol{\eta})}{g(X_i|X_{i-1}^p; \vartheta_0, \boldsymbol{\eta}_0)}\Bigg|_{\eta_j = \eta_{j,0}}, \\
U_i^{\eta_j} &= U_i^{\eta_j}(\boldsymbol{\eta}_0), \\
Y_i(\vartheta_0, \boldsymbol{\eta}) &= \frac{\frac{\partial}{\partial \vartheta} g(X_i|X_{i-1}^p; \vartheta, \boldsymbol{\eta})}{g(X_i|X_{i-1}^p; \vartheta_0, \boldsymbol{\eta}_0)}\Bigg|_{\vartheta = \vartheta_0}, \\
Y_i &= Y_i(\vartheta_0, \boldsymbol{\eta}_0).
\end{aligned} \tag{2.2.6}$$

Assumption 2.4. There exists a nonnegative function K such that $EK(X_1^{p+1}) < \infty$ and such that for all $\boldsymbol{\eta} \in \boldsymbol{H}$, $\vartheta \in \Theta$ and $x_1^{p+1} \in \mathbb{R}^{p+1}$,

$$|Y_1(\vartheta, \boldsymbol{\eta})|^3 \leq K(x_1^{p+1}), \ |Z_1(\vartheta)|^3 \leq K(x_1^{p+1}), \ |U_1^{\eta_j}(\boldsymbol{\eta})|^3 \leq K(x_1^{p+1})$$

for all $j = 1, \ldots, d$.

Assumption 2.5. We have that

$$R_n = 2\{l_n(\widehat{\alpha}, \widehat{\vartheta}_1, \widehat{\vartheta}_2, \widehat{\boldsymbol{\eta}}) - l_n(1/2, \widehat{\vartheta}_0, \widehat{\vartheta}_0, \widehat{\boldsymbol{\eta}}_0)\} = O_P(1). \tag{2.2.7}$$

This assumption is only used to show that $p(\widehat{\alpha}^*) = O_P(1)$, where $\widehat{\alpha}^*$ is the modified likelihood estimator for α. Its validity under general assumptions follows from the results in Cho and White (2007).

Assumptions 2.1 – 2.5 are further discussed below Theorem 2.2.

Theorem 2.1. *Suppose that Assumptions 2.1 – 2.5 are satisfied. In case of a single state (i.e. no switching regime), we have that*

(i) $\widehat{\vartheta}_0 - \vartheta_0 = o_P(1)$, $\widehat{\boldsymbol{\eta}}_0 - \boldsymbol{\eta}_0 = o_P(1)$ *and*

(ii) $\widehat{\vartheta}_1^* - \vartheta_0 = o_P(1)$, $\widehat{\vartheta}_2^* - \vartheta_0 = o_P(1)$, $\widehat{\boldsymbol{\eta}}^* - \boldsymbol{\eta}_0 = o_P(1)$.

Remark 2.1. Under the hypothesis of no regime switching, both estimators $\widehat{\vartheta}_i^*$ are consistent for ϑ_0. This is due to the penalty term $p(\alpha)$ in (2.2.5): The estimator $\widehat{\alpha}^*$ is forced to be bounded away from 0 and 1, so that both $\widehat{\vartheta}_i^*$ need to be consistent. This is not true for the quasi MLEs $\widehat{\vartheta}_i$.

2.3 Feasible quasi-likelihood based tests for regime switching

2.3.1 The modified quasi-likelihood ratio test

If $(1-\alpha, \alpha)$ denotes the stationary distribution of $(S_k)_k$, then the hypothesis of no regime switch is equivalent to

$$H : \alpha(1-\alpha)(\vartheta_1 - \vartheta_2) = 0.$$

We propose to test H via the *modified quasi likelihood ratio test* (MQLRT) statistic

$$M_n = 2\{pl_n(\widehat{\alpha}^*, \widehat{\vartheta}_1^*, \widehat{\vartheta}_2^*, \widehat{\boldsymbol{\eta}}^*) - pl_n(1/2, \widehat{\vartheta}_0, \widehat{\vartheta}_0, \widehat{\boldsymbol{\eta}}_0)\}. \tag{2.3.1}$$

In order to derive the asymptotic distribution of the MQLRT, we need the following additional assumptions, which are further discussed below.

Assumption 2.6. The covariance matrix of $(U_1^{\eta_1}, \ldots, U_1^{\eta_d}, Y_1, Z_1)$ is positive definite.

Assumption 2.7. The processes

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{U_i^{\eta_j}(\boldsymbol{\eta}) - U_i^{\eta_j}(\eta_1, \ldots, \eta_{k-1}, \eta_{k,0}, \eta_{k+1}, \ldots, \eta_d)}{\eta_k - \eta_{k,0}}, \quad 1 \le j,k \le d,$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Y_i(\vartheta, \boldsymbol{\eta}) - Y_i(\vartheta, \eta_1, \ldots, \eta_{k-1}, \eta_{k,0}, \eta_{k+1}, \ldots, \eta_d)}{\eta_k - \eta_{k,0}}, \quad 1 \le k \le d,$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{Z_i(\vartheta)-Z_i}{\vartheta-\vartheta_0}$$

are tight.

Theorem 2.2. *Under the null hypothesis H of no regime switching, if Assumptions 2.1 – 2.7 are satisfied we have that*

$$M_n \xrightarrow{d} \frac{1}{2}\,\chi_0^2 + \frac{1}{2}\,\chi_1^2,$$

where χ_p^2 denotes the χ^2-distribution with $p > 0$ degrees of freedom, χ_0^2 is the point mass at 0, and $\xrightarrow{d}$ denotes convergence in distribution.

Remark 2.2. As desired, the asymptotic distribution of M_n is easy to handle and does not depend on the underlying parametric model, the actual true parameter values or the choice of the compact set Θ (as long as it contains the true value). This is in contrast to the asymptotic distribution of a quasi LRT under the hypothesis based on the quasi log-likelihood function (1.5.4), cf. Cho and White (2007). Note that Assumption 2.6 corresponds to the case of a non-zero second order derivative as discussed in Cho and White (2007). In case of a zero second-order derivative, which arises in particular in linear switching AR models with possibly switching intercept under the alternative, normal innovations and structural scale parameter (see Example 2.1 below), the asymptotic distribution in Theorem 2.2 does no longer hold true. We will deal with this case in the following chapter. This is also known for normal mixtures, see e.g. Chen and Chen (2003).

Example 2.1 *(continued).*

First, consider Assumption 2.1. Consider the AR(p) process $(X_k)_k$ defined by

$$X_t = \zeta + \sum_{j=1}^{p}\phi_j X_{t-j} + \sigma\epsilon_t,$$

where $(\epsilon_k)_k$ are i.i.d. random variables with $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = 1$. We denote by $\rho(\boldsymbol{A})$ the spectral radius of the matrix

$$\boldsymbol{A} = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \cdots & \phi_p \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \ddots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \in \mathbb{R}^{p\times p}.$$

Under conditions which are fulfilled for the t- as well as for the normal distribution, the process $(X_k)_k$ is geometrically ergodic if $\rho(A) < 1$, see e.g. Lu (1998, Thm. 2). This is also the condition for stationarity.

Now, consider the identifiability conditions (Assumptions 2.2 and 2.6). Suppose that the innovations are real-valued with continuous density $f > 0$ w.r.t. Lebesgue measure and let $f(x;\mu,\sigma) = f\big((x-\mu)/\sigma\big)/\sigma$ denote the corresponding location-scale family, so that the conditional density of X_1 is given by

$$g(x_1|x_0^p;\zeta,\phi_1,\ldots,\phi_p,\sigma) = \frac{1}{\sigma} f\Big(\frac{x_1-\zeta-\phi_1 x_0-\ldots-\phi_p x_{1-p}}{\sigma}\Big). \tag{2.3.2}$$

Then we have the following lemma.

Lemma 2.1. (*i*). *If the parameter* $(\alpha,\mu_1,\mu_2,\sigma)$ *in a two-component location mixture* $(1-\alpha)f(x;\mu_1,\sigma)+\alpha f(x;\mu_2,\sigma)$ *is identifiable (except for label switching), then Assumption 2.2 holds for model* (2.2.1) *and* (2.2.2).
(*ii*). *If the parameter* $(\alpha,\mu,\sigma_1,\sigma_2)$ *in a two-component scale mixture* $(1-\alpha)f(x;\mu,\sigma_1)+\alpha f(x;\mu,\sigma_2)$ *is identifiable (except for label switching), then Assumption 2.2 holds for model* (2.2.3).

The simple proof is omitted. Since general finite mixtures of normal and t-distributions (even with variable degrees of freedom) are identifiable (cf. Holzmann et al. 2006), Assumption 2.2 will also be satisfied.

Lemma 2.2. *Suppose that for any* (μ,σ) *and* $a_1,a_2,a_3\in\mathbb{R}$,

$$a_1\frac{\partial f(x;\mu,\sigma)}{\partial\mu}+a_2\frac{\partial^2 f(x;\mu,\sigma)}{\partial^2\mu}+a_3\frac{\partial f(x;\mu,\sigma)}{\partial\sigma}=0 \quad \textit{for Lebesgue-a.e. } x \tag{2.3.3}$$

entails $a_1=a_2=a_3=0$. *Then Assumption 2.6 is satisfied for the models* (2.2.1) *and* (2.2.2).

The next lemma shows that the condition of Lemma 2.2 and hence Assumption 2.6 is indeed satisfied for the t-distribution. Since $\sigma\frac{\partial^2 f(x;\mu,\sigma)}{\partial^2\mu}=\frac{\partial f(x;\mu,\sigma)}{\partial\sigma}$ holds for the normal distribution, condition (2.3.3) is not fulfilled. Hence, for model (2.2.1) the MQLRT for testing for homogeneity does not admit the simple asymptotic distribution given in Theorem 2.2 in case of a variable scale parameter (it does for fixed scale parameter). Therefore, we treat this case in Chapter 3 separately and give a feasible method for testing for homogeneity. Even if (2.3.3) is not satisfied for the normal distribution, Lemma 2.4 shows that Assumption 2.6 is satisfied for model (2.2.2).

Lemma 2.3. *For a fixed* ν, *let* $f(x)=\Gamma\left(\frac{\nu+1}{2}\right)\left(\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\left(1+\frac{x^2}{\nu}\right)^{(\nu+1)/2}\right)^{-1}$ *be the density of the* t*-distribution with* ν *degrees of freedom. Then for the associated location-scale family* $f(x;\mu,\sigma)$, *for any* (μ,σ) *and* $a_1,a_2,a_3,b_1,b_2,b_3\in\mathbb{R}$,

(*i*)

$$a_1\frac{\partial f(x;\mu,\sigma)}{\partial\mu}+a_2\frac{\partial^2 f(x;\mu,\sigma)}{\partial^2\mu}+a_3\frac{\partial f(x;\mu,\sigma)}{\partial\sigma}=0 \quad \textit{for Leb.-a.e. } x \tag{2.3.4}$$

entails $a_1 = a_2 = a_3 = 0$.

(ii)

$$b_1 \frac{\partial f(x;\mu,\sigma)}{\partial \mu} + b_2 \frac{\partial f(x;\mu,\sigma)}{\partial \sigma} + b_3 \frac{\partial^2 f(x;\mu,\sigma)}{\partial^2 \sigma} = 0 \quad \textit{for Leb.-a.e. } x \tag{2.3.5}$$

entails $b_1 = b_2 = b_3 = 0$.

Lemma 2.4. *For the normal distribution, Assumption 2.6 is satisfied for the model* (2.2.2).

Assumptions 2.3 and 2.4 are satisfied for the t- and the normal distribution since Θ and $\boldsymbol{H}$ are assumed to be compact. Since $(X_k)_k$ is geometrically ergodic and therefore strongly mixing with exponentially decaying coefficients (cf. Bradley, 2005), Assumption 2.7 can be verified as in Fu, Chen and Li (2008) using the same modification as Dannemann (2009). Assumption 2.5 follows from the results in Cho and White (2007).

Lemma 2.5. *Suppose that for any* (μ, σ) *and* $a_1, a_2, a_3 \in \mathbb{R}$,

$$a_1 \frac{\partial f(x;\mu,\sigma)}{\partial \mu} + a_2 \frac{\partial f(x;\mu,\sigma)}{\partial \sigma} + a_3 \frac{\partial^2 f(x;\mu,\sigma)}{\partial^2 \sigma} = 0 \quad \textit{for Lebesgue-a.e. } x \tag{2.3.6}$$

entails $a_1 = a_2 = a_3 = 0$. *Then Assumption 2.6 is satisfied for the model* (2.2.3).

The following lemma shows that the condition of Lemma 2.5 holds for the normal distribution. For the t-distribution, this condition is satisfied by Lemma 2.3 *(ii)*.

Lemma 2.6. *Let* $f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ *be the pdf of a normally distributed random variable with expectation* μ *and standard deviation* $\sigma > 0$. *If*

$$a_1 \frac{\partial f(x;\mu,\sigma)}{\partial \mu} + a_2 \frac{\partial f(x;\mu,\sigma)}{\partial \sigma} + a_3 \frac{\partial^2 f(x;\mu,\sigma)}{\partial^2 \sigma} = 0 \quad \textit{for Lebesgue-a.e. } x \tag{2.3.7}$$

for any (μ, σ) *and* $a_1, a_2, a_3 \in \mathbb{R}$, *then* $a_1 = a_2 = a_3 = 0$.

Example 2.2 *(continued).*

First, consider Assumption 2.1. Consider the pure ARCH(p) process for which for every t,

$$X_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \vartheta + \sum_{j=1}^{p} \phi_j X_{t-j}^2,$$

where $(\epsilon_k)_k$ are i.i.d. random variables with $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = 1$, holds. Under conditions which are fulfilled for the t- as well as for the normal distribution, the process

$(X_k)_k$ is geometrically ergodic if

$$\sum_{j=1}^{p} \phi_j < 1,$$

see e.g. Lu (1998, Cor. 1). This is also the condition for the strictly stationarity, see e.g. Fan and Yao (2003).

Again we concentrate on the identifiability Assumptions 2.2 and 2.6. Suppose that the innovations are real-valued with continuous density $f > 0$ w.r.t. Lebesgue measure, and let $f(x; \sigma) = f(x/\sigma)/\sigma$ denote the corresponding scale family, so that conditional density of X_1 is given by

$$g(x_1|x_0^p; \vartheta, \boldsymbol{\phi}) = f\big(x_1; \sigma(\vartheta, \boldsymbol{\phi}, x_0^p)\big), \qquad \sigma^2(\vartheta, \boldsymbol{\phi}, x_0^p) = \vartheta + \phi_1 x_0^2 + \ldots + \phi_d x_{1-p}^2, \tag{2.3.8}$$

where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^T$. Then we have the following

Lemma 2.7. *If the parameter $(\alpha, \sigma_1, \sigma_2)$ in a two-component scale mixture $(1-\alpha)f(x; \sigma_1) + \alpha f(x; \sigma_2)$ is identifiable (except for label switching), then Assumption 2.2 holds for model* (2.2.4).

This is satisfied by normal and t-distributions.

Lemma 2.8. *Suppose that for any $\sigma > 0$ and $a_1, a_2 \in \mathbb{R}$,*

$$a_1 \frac{\partial f(x; \mu, \sigma)}{\partial \sigma} + a_2 \frac{\partial^2 f(x; \mu, \sigma)}{\partial^2 \sigma} = 0 \quad \textit{for Lebesgue-a.e. } x \tag{2.3.9}$$

entails $a_1 = a_2 = 0$. Then Assumption 2.6 is satisfied for the model (2.2.4).

For the normal distribution, this is implied by strong identifiability of the $N(\mu_0, \sigma^2)$ (for fixed μ_0 as a scale-family, see Chen 1995) or by Lemma 2.6 and for the t-distribution, it follows from Lemma 2.3 *(ii)*.

2.3.2 The EM-test

The MQLRT is simple to compute and has a tractable asymptotic distribution as specified in Theorem 2.2. The proof of Theorem 2.2 shows that the asymptotic distribution is dominated by $\alpha = 1/2$, in other words, the same asymptotic distribution arises if under the alternative, $\alpha = 1/2$ is fixed. Since this evidently only holds asymptotically, the test will sometimes be anticonservative in finite samples.

Therefore, in the context of finite mixtures Chen and Li (2009) introduced the EM-test. The idea is not to maximize the quasi-likelihood (2.2.5) over all α, instead, one starts with a finite set of initial values, say $\mathcal{J} = \{\alpha_1, \ldots, \alpha_J\}$, for α and proceeds from these by a finite number of steps of the EM algorithm. If one of the initial values for α is $1/2$, the asymptotic distribution will be the same as that of the MQLRT.

We now proceed to describe the EM-test, which is most conveniently accomplished in form of the following algorithm.

Step 0. Choose the initial values $0 < \alpha_1 < \alpha_2 < \ldots < \alpha_J = 0.5$. Compute

$$(\tilde{\vartheta}_0, \tilde{\boldsymbol{\eta}}_0) = \arg\max_{\vartheta,\boldsymbol{\eta}} pl_n(0.5, \vartheta, \vartheta, \boldsymbol{\eta}).$$

Put $j = 1$ and $k = 0$.

Step 1. Put $\alpha_j^{(k)} = \alpha_j$.

Step 2. Compute

$$(\vartheta_{1j}^{(k)}, \vartheta_{2j}^{(k)}, \boldsymbol{\eta}_j^{(k)}) = \arg\max_{\vartheta_1,\vartheta_2,\boldsymbol{\eta}} pl_n(\alpha_j^{(k)}, \vartheta_1, \vartheta_2, \boldsymbol{\eta})$$

and

$$M_n^{(k)}(\alpha_j) = 2\left(pl_n(\alpha_j^{(k)}, \vartheta_{1j}^{(k)}, \vartheta_{2j}^{(k)}, \boldsymbol{\eta}_j^{(k)}) - pl_n(0.5, \tilde{\vartheta}_0, \tilde{\vartheta}_0, \tilde{\boldsymbol{\eta}}_0)\right).$$

Step 3. Compute for $i = 1, \ldots, n$ the weights

$$w_{ij}^{(k)} = \frac{\alpha_j^{(k)} g(X_i|X_{i-1}^p; \vartheta_{2j}^{(k)}, \boldsymbol{\eta}_j^{(k)})}{(1-\alpha_j^{(k)})\, g(X_i|X_{i-1}^p; \vartheta_{1j}^{(k)}, \boldsymbol{\eta}_j^{(k)}) + \alpha_j^{(k)} g(X_i|X_{i-1}^p; \vartheta_{2j}^{(k)}, \boldsymbol{\eta}_j^{(k)})}.$$

Compute the estimators

$$\begin{aligned}
\alpha_j^{(k+1)} &= \arg\max_{\alpha}\left(\left(n - \sum_{i=1}^n w_{ij}^{(k)}\right)\log(1-\alpha) + \sum_{i=1}^n w_{ij}^{(k)}\log(\alpha) + p(\alpha)\right) \\
\vartheta_{1j}^{(k+1)} &= \arg\max_{\vartheta_1}\left(\sum_{i=1}^n (1-w_{ij}^{(k)})\log g(X_i|X_{i-1}^p; \vartheta_1, \boldsymbol{\eta}_j^{(k)})\right) \\
\vartheta_{2j}^{(k+1)} &= \arg\max_{\vartheta_2}\left(\sum_{i=1}^n w_{ij}^{(k)}\log g(X_i|X_{i-1}^p; \vartheta_2, \boldsymbol{\eta}_j^{(k)})\right)
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\eta}_j^{(k+1)} &= \arg\max_{\boldsymbol{\phi}}\left(\sum_{i=1}^n (1-w_{ij}^{(k)})\log g(X_i|X_{i-1}^p; \vartheta_{1j}^{(k+1)}, \boldsymbol{\eta})\right. \\
&\qquad \left. + \sum_{i=1}^n w_{ij}^{(k)}\log g(X_i|X_{i-1}^p; \vartheta_{2j}^{(k+1)}, \boldsymbol{\eta})\right).
\end{aligned}$$

Compute

$$M_n^{(k+1)}(\alpha_j) = 2\left\{pl_n(\alpha_j^{(k+1)}, \vartheta_{1j}^{(k+1)}, \vartheta_{2j}^{(k+1)}, \boldsymbol{\eta}_j^{(k+1)}) - pl_n(0.5, \tilde{\vartheta}_0, \tilde{\vartheta}_0, \tilde{\boldsymbol{\eta}}_0)\right\},$$

put $k = k + 1$ and repeat Step 3 for a fixed number of iterations K.

Step 4. Put $j = j + 1$, $k = 0$ and go to Step 1, until $j = J$.

Step 5. Compute the test statistic

$$EM_n^{(K)} = \max_{j=1,\ldots,J} M_n^{(K)}(\alpha_j).$$

The following theorem is a direct consequence of the proof of Theorem 2.2.

Theorem 2.3. *Under the assumptions of Theorem 2.2, if one of the initial values for α in the above algorithm is equal to $1/2$, we have*

$$EM_n^{(K)} \xrightarrow{d} \frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2. \tag{2.3.10}$$

Remark 2.3. a. Our simulations (see Section 2.4) show that only in case of a variable scale parameter in AR-models with t-distributed innovations, the EM-test actually performs better than the MQLRT, otherwise both tests perform virtually identical. Furthermore, very few iterations $K = 0, 1, 2$ suffice to capture the power of the EM-test.
b. In cases where Theorem 2.2 does not apply (e.g. normal location mixtures with variable, equal scale parameter), the asymptotic distribution of the EM-test is still accessible (though different from that in Theorem 2.3), cf. Chapter 3 or Chen and Li (2009).
c. In the construction of the EM-test, we actually use an ECM algorithm (Meng and Rubin, 1993) since the EM algorithm would require joint maximization to obtain the update $(\vartheta_{1j}^{(k+1)}, \vartheta_{2j}^{(k+1)}, \boldsymbol{\eta}_j^{(k+1)})$. If $\boldsymbol{\eta}$ is high-dimensional, this could be further refined by maximizing successively over the components of $\boldsymbol{\eta}$.

2.4 Simulations

Here we present some of the results of an extensive simulation study of the tests proposed in the two previous sections. In the simulations we choose the second penalty function in (1.5.6) with $C = 1$, if not stated otherwise.

2.4.1 Simulated sizes

In this section we simulate the size of the MQLRT and the EM-test in several settings. As suggested in Chen and Li (2009) we choose $\mathcal{J} = \{0.1, 0.3, 0.5\}$ as initial values for the EM-test.

a. Switching Autoregression witch switching intercept and $N(0, 1)$-distributed innovations.

Data-generating process (DGP): $X_t = 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0, 1)$.

Model: $X_t = \zeta_{S_t} + \phi X_{t-1} + \epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$.
The results for various sample sizes are contained in Table 2.1. Both tests are somewhat conservative and have almost identical levels.

Table 2.1: DGP: $X_t = 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, Model: $X_t = \zeta_{S_t} + \phi X_{t-1} + \epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$; number of replications: 20,000.

Sample Size	Nominal Levels (%)	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$MQLRT$
$n = 100$	10%	7.3	7.3	7.3	7.4
	5%	3.6	3.6	3.6	3.8
	1%	0.7	0.7	0.7	0.8
$n = 200$	10%	8.1	8.1	8.1	8.1
	5%	3.9	3.9	3.9	4.0
	1%	0.7	0.7	0.7	0.8
$n = 500$	10%	8.7	8.7	8.7	8.7
	5%	4.4	4.4	4.4	4.4
	1%	1.0	1.0	1.0	1.0

b. Switching Autoregression with switching intercept with t-distributed innovations and variable scale.
DGP: $X_t = 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} t(5)$.
Model: $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\,\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} t(5)$.
Fitting a model with switching regime to data arising from a model without switching regime tends to result in a lower fitted standard deviation σ. Therefore following Chen and Li (2009) we add the penalty function

$$\tilde{p}(\sigma) = C^* \log(1 + |\sigma - \hat{\sigma}_0|) \tag{2.4.1}$$

to pl_n in (2.2.5), which prevents under-estimation of σ. Here C^* is a non-positive constant and $\hat{\sigma}_0$ is the MLE for σ under the null model (i.e. no switching regime). Due to Theorem 2.1 and the fact that $\tilde{p}(\sigma)$ is a continuous function this penalty is asymptotically negligible, cf. Chen and Li (2009). For our simulation study we choose $C = 3$ in (1.5.6) and $C^* = -1$ in (2.4.1). In the last column of Table 2.2 we report the results of our simulation study when $C^* = 0$ (i.e. the penalized likelihood function is given by (2.2.5)). The results are given in Table 2.2. Without penalty, the MQLRT (and also EM-test if slightly less, results not shown) are anticonservative, particularly for the sample size $n = 100$. If we add the penalty on σ, both the EM-test and the MQLRT keep the nominal levels quite well, the EM-test being slightly better.

Table 2.2: DGP: $X_t = 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} t(5)$, Model: $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} t(5)$; number of replications: 20,000.

Sample Size	Nominal Levels (%)	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$MQLRT$	$MQLRT^*$
$n = 100$	10%	11.1	11.3	11.4	11.8	14.6
	5%	6.0	6.2	6.3	6.6	8.4
	1%	1.4	1.5	1.5	1.6	2.2
$n = 200$	10%	10.2	10.3	10.3	10.6	12.6
	5%	5.3	5.3	5.3	5.5	6.4
	1%	1.2	1.2	1.2	1.3	1.6
$n = 500$	10%	10.0	10.0	10.1	10.2	11.5
	5%	5.3	5.3	5.3	5.5	6.4
	1%	1.2	1.2	1.2	1.3	1.6

c. Switching ARCH

DGP 1: $X_t = \sigma_t\epsilon_t$; $\sigma_t^2 = 1 + 0.5X_{t-1}^2 + 0.3X_{t-2}^2$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$.
Model 1: $X_t = \sigma_t\epsilon_t$; $\sigma_t^2 = \vartheta_{S_t} + \phi_1 X_{t-1}^2 + \phi_2 X_{t-2}^2$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$.
The results are in Table 2.3 for DGP 1 and Model 1. For DGP 1, the tests are slightly conservative.

Table 2.3: DGP: $X_t = \sigma_t\epsilon_t$; $\sigma_t^2 = 1 + 0.5X_{t-1}^2 + 0.3X_{t-2}^2$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, Model: $X_t = \sigma_t\epsilon_t$; $\sigma_t^2 = \vartheta_{S_t} + \phi_1 X_{t-1}^2 + \phi_2 X_{t-2}^2$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$; number of replications: 20,000.

Sample Size	Nominal Levels (%)	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$MQLRT$
$n = 100$	10%	8.0	8.0	8.1	8.4
	5%	4.4	4.4	4.5	4.7
	1%	0.9	0.9	0.9	1.1
$n = 200$	10%	8.3	8.3	8.3	8.6
	5%	4.5	4.5	4.5	4.8
	1%	1.0	1.0	1.0	1.1
$n = 500$	10%	8.8	8.8	8.8	9.1
	5%	4.6	4.6	4.6	4.9
	1%	1.1	1.1	1.1	1.1

DGP 2: $X_t = \sigma_t\epsilon_t$; $\sigma_t^2 = 1 + 0.5X_{t-1}^2$, where $\epsilon_t \overset{iid}{\sim} t(5)$.
Model 2: $X_t = \sigma_t\epsilon_t$; $\sigma_t^2 = \vartheta_{S_t} + \phi X_{t-1}^2$ with $\epsilon_t \overset{iid}{\sim} t(5)$.
The results are given in Table 2.4 for DGP 2 and Model 2. For DGP 2, the tests keep the nominal level almost exactly.

Table 2.4: DGP: $X_t = \sigma_t \epsilon_t$; $\sigma_t^2 = 1 + 0.5X_{t-1}^2$, where $\epsilon_t \overset{iid}{\sim} t(5)$. Model: $X_t = \sigma_t \epsilon_t$; $\sigma_t^2 = \vartheta_{S_t} + \phi X_{t-1}^2$ with $\epsilon_t \overset{iid}{\sim} t(5)$; number of replications: 20,000.

Sample Size	Nominal Levels (%)	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$MQLRT$
$n = 100$	10%	9.6	9.7	9.8	10.1
	5%	5.1	5.2	5.2	5.5
	1%	1.2	1.2	1.2	1.3
$n = 200$	10%	9.5	9.5	9.6	9.9
	5%	5.1	5.2	5.2	5.5
	1%	1.2	1.3	1.3	1.4
$n = 500$	10%	9.7	9.7	9.7	9.9
	5%	5.3	5.3	5.3	5.5
	1%	1.2	1.2	1.2	1.3

d. Switching Autoregression with normally distributed innovations with switching scale.
DGP: $X_t = 0.2X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$.
Model: $X_t = \zeta + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$.
The results for various sample sizes are contained in Table 2.5. Both tests have almost identical levels.

Table 2.5: DGP: $X_t = 0.2X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, Model: $X_t = \zeta + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$; number of replications: 20,000.

Sample Size	Nominal Levels (%)	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$MQLRT$
$n = 200$	10%	11.0	11.1	11.1	11.4
	5%	6.3	6.4	6.4	6.7
	1%	1.9	1.9	2.0	2.1
$n = 500$	10%	9.6	9.7	9.7	9.8
	5%	5.1	5.1	5.1	5.3
	1%	1.1	1.1	1.1	1.2

e. Switching Autoregression with t-distributed innovations with switching scale.
DGP: $X_t = -0.3X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} t(9)$.
Model: $X_t = \zeta + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} t(9)$.
The results for various sample sizes are contained in Table 2.6. Both tests are somewhat anticonservative and have almost identical levels.

Table 2.6: DGP: $X_t = -0.3X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} t(9)$, Model: $X_t = \zeta + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} t(9)$; number of replications: 20,000.

Sample Size	Nominal Levels (%)	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$MQLRT$
$n = 200$	10%	11.4	11.6	11.6	11.9
	5%	6.9	7.0	7.0	7.4
	1%	2.2	2.3	2.4	2.7
$n = 500$	10%	10.2	10.2	10.2	10.5
	5%	5.4	5.5	5.5	5.7
	1%	1.3	1.3	1.4	1.5

f. Switching Autoregression with normally distributed innovations with switching autoregressive parameter.

DGP: $X_t = 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$.

Model 1: $X_t = \zeta + \phi_{S_t}X_{t-1} + \sigma\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$.

Model 2: $X_t = \phi_{S_t}X_{t-1} + \sigma\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$. This is the model discussed in Lange and Rahbek (2009).

The results for various sample sizes are contained in Table 2.7 (for Model 1) and 2.8 (for Model 2). Both tests have almost identical levels and are somewhat conservative.

Table 2.7: DGP: $X_t = 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, Model: $X_t = \zeta + \phi_{S_t}X_{t-1} + \sigma\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$; number of replications: 20,000.

Sample Size	Nominal Levels (%)	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$MQLRT$
$n = 200$	10%	7.9	8.0	8.0	8.6
	5%	4.2	4.2	4.2	4.8
	1%	1.0	1.0	1.0	1.2
$n = 500$	10%	8.9	8.9	8.9	9.3
	5%	4.8	4.8	4.8	5.3
	1%	1.0	1.0	1.0	1.3

Table 2.8: DGP: $X_t = 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, Model: $X_t = \phi_{S_t}X_{t-1} + \sigma\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$; number of replications: 20,000.

Sample Size	Nominal Levels (%)	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$MQLRT$
$n = 200$	10%	8.5	8.6	8.6	8.9
	5%	4.6	4.7	4.7	5.0
	1%	1.1	1.2	1.2	1.3
$n = 500$	10%	9.0	9.0	9.0	9.3
	5%	4.8	4.8	4.8	5.0
	1%	1.1	1.1	1.1	1.2

g. Switching Autoregression with t-distributed innovations with switching autoregressive parameter.

DGP: $X_t = 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} t(9)$.
Model: $X_t = \zeta + \phi_{S_t}X_{t-1} + \sigma\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} t(9)$.
The results for various sample sizes are contained in Table 2.9.

Table 2.9: DGP: $X_t = 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} t(9)$, Model: $X_t = \zeta + \phi_{S_t}X_{t-1} + \sigma\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} t(9)$; number of replications: 20,000.

Sample Size	Nominal Levels (%)	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$MQLRT$
$n = 200$	10%	9.8	9.9	9.9	10.7
	5%	5.0	5.0	5.1	5.9
	1%	1.0	1.0	1.0	1.4
$n = 500$	10%	9.5	9.6	9.6	10.3
	5%	5.0	5.0	5.1	5.7
	1%	1.1	1.1	1.1	1.4

Summarizing, in most scenarios, both the MQLRT and the EM-test have appropriate sizes. Further, the EM-test outperforms the MQLRT only for a switching AR(p)-model (in the intercept as well as in the autoregressive parameter) with variable scale (and t-innovations), otherwise, the computationally simpler MQLRT should be preferred.

2.4.2 Power comparison of several tests

Here we conduct a power comparison between the MQLRT, the EM-test and the QLRT by Cho and White (2007). In order to properly estimate the power we used simulated critical values in all scenarios for each of the tests. More precisely, for given alternative, we simulate

the critical values of the tests from the distribution (without switching regime) which is fitted to a large sample (sample size $n = 10,000$) from the alternative by (conditional) maximum likelihood. Note the analogy to a corresponding bootstrap procedure.
DGP 1: $X_t = (-1)^{S_t}\zeta + 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, with $a_{12} \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, $a_{12} = a_{21}$ and $\zeta \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$.
Model 1: $X_t = \zeta_{S_t} + \phi X_{t-1} + \epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$.
Note that in contrast to the simulations in Cho and White (2007, Section 3, Table 3), we fix the scale at $\sigma = 1$. The results can be found in Table 2.10. The EM-test and the MQLRT perform almost identically, which is not surprising since one of the starting values in the EM-test is $\alpha = 0.5$, the true parameter value. The MQLRT and the QLRT have comparable power properties, indeed, the MQLRT has even somewhat higher power in these scenarios. We also computed the Bera and Jarque statistic (BJ) and Neyman and Scott's $C(\alpha)$, however, as in Cho and White (2007) these have virtually no power against the alternatives under consideration, therefore we do not report the results.

Table 2.10: Nominal level: 0.05; DGP: $X_t = (-1)^{S_t}\zeta + 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, sample size: 100, number of replications: 5,000, Model: $X_t = \zeta_{S_t} + \phi X_{t-1} + \epsilon_t$, with $\epsilon_t \overset{iid}{\sim} N(0,1)$.

a_{12}	ζ	0.1	0.2	0.3	0.4	0.5
0.1	QLRT	5.8	7.8	12.5	19.8	30.4
	$EM_n^{(0)}$	5.7	8.2	14.3	23.5	36.7
	$EM_n^{(1)}$	5.7	8.2	14.3	23.5	36.7
	MQLRT	5.7	8.2	14.2	23.5	36.8
0.3	QLRT	6.0	8.0	14.0	24.1	40.9
	$EM_n^{(0)}$	6.4	8.9	16.4	28.2	48.8
	$EM_n^{(1)}$	6.4	8.9	16.4	28.2	48.8
	MQLRT	6.4	9.0	16.3	28.2	48.7
0.5	QLRT	6.0	8.0	13.6	24.1	42.5
	$EM_n^{(0)}$	6.1	8.4	16.1	29.0	51.2
	$EM_n^{(1)}$	6.1	8.4	16.1	29.0	51.2
	MQLRT	6.0	8.5	16.0	29.0	51.1
0.7	QLRT	5.4	7.5	13.6	22.7	40.6
	$EM_n^{(0)}$	5.9	8.2	15.6	28.3	47.3
	$EM_n^{(1)}$	5.9	8.2	15.6	28.3	47.3
	MQLRT	6.0	8.1	15.6	28.2	47.3
0.9	QLRT	5.7	8.1	12.7	21.4	33.9
	$EM_n^{(0)}$	5.8	9.2	15.5	27.0	45.1
	$EM_n^{(1)}$	5.8	9.3	15.6	27.0	45.1
	MQLRT	5.8	9.2	15.5	27.1	45.2

DGP 2: $X_t = (-1)^{S_t}\zeta + 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, $\zeta \in \{0.1, 0.2, \ldots, 0.8\}$, $\alpha \in \{0.1, 0.2, 0.3, 0.4\}$, various combinations of a_{12} and a_{21} (leading to the proper values of α).
Model 2: $X_t = \zeta_{S_t} + \phi X_{t-1} + \epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$. Again, we fix the scale at $\sigma = 1$. The results can be found in Table 2.11. Again, the EM-test and the MQLRT perform almost identically, therefore we do not display the results of the EM-test. The QLRT and MQLRT still perform comparably and we observe that especially for small values of α, the MQLRT no longer has higher power.

Table 2.11: Nominal level: 5%; DGP: $X_t = (-1)^{S_t}\zeta + 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, sample size: 100, number of replications: 5,000, Model: $X_t = \zeta_{S_t} + \phi X_{t-1} + \epsilon_t$, with $\epsilon_t \overset{iid}{\sim} N(0,1)$. Let $\alpha = a_{12}/(a_{12} + a_{21})$ and $(1-\alpha, \alpha)$ be the stationary distribution of the hidden Markov Chain $(S_k)_k$.

a_{12}	a_{21}	α	ζ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
0.1	0.9	0.1	QLRT	5.8	6.6	8.0	10.8	15.7	22.6	33.8	46.8
			MQLRT	5.0	6.1	7.7	10.7	15.6	22.8	33.0	44.4
0.05	0.45		QLRT	5.7	7.1	8.8	11.4	16.1	23.4	32.5	42.9
			MQLRT	4.9	6.6	8.3	11.6	16.6	22.9	32.0	42.1
0.2	0.8	0.2	QLRT	5.8	7.6	10.6	17.7	27.1	43.3	61.6	71.9
			MQLRT	5.3	7.1	11.3	18.3	28.4	45.8	64.1	72.9
0.1	0.4		QLRT	5.8	7.4	10.7	17.4	26.4	42.5	60.5	70.1
			MQLRT	5.3	6.9	10.9	18.0	28.0	44.9	62.7	71.6
0.05	0.2		QLRT	5.9	7.4	10.2	15.8	23.2	32.8	44.9	57.9
			MQLRT	5.3	7.1	10.5	15.9	23.2	34.9	47.1	59.4
0.3	0.7	0.3	QLRT	5.9	8.2	12.9	22.1	37.0	56.9	77.3	90.9
			MQLRT	5.5	7.6	13.1	23.3	40.0	61.2	81.0	93.2
0.15	0.35		QLRT	5.8	8.7	13.2	22.2	35.7	53.3	72.0	86.6
			MQLRT	5.4	7.9	13.5	23.0	38.4	56.6	75.6	88.4
0.4	0.6	0.4	QLRT	6.1	9.0	15.3	26.2	43.3	64.0	83.3	94.6
			MQLRT	5.5	8.8	15.4	28.5	46.9	68.5	86.6	96.4
0.2	0.3		QLRT	5.9	8.5	15.1	26.0	42.3	61.2	80.5	92.6
			MQLRT	5.6	8.7	15.2	27.4	45.2	66.2	84.0	94.9
0.1	0.15		QLRT	6.0	8.8	14.0	22.7	36.0	51.2	67.6	80.7
			MQLRT	5.6	8.7	14.4	24.0	38.3	54.7	71.3	84.4

DGP 3: $X_t = (-1)^{S_t}\zeta + 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} t(5)$, various values of ζ, α, a_{12} and a_{21}.
Model 3: $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} t(5)$.
Note that we estimate the scale σ of the error distribution. The results can be found in Table 2.12. Here, the power of the EM-test increases somewhat when using 2 instead of 0 or 1 iterations and approaches that of the MQLRT.

Table 2.12: Nominal level: 5%; DGP: $X_t = (-1)^{S_t}\zeta + 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} t(5)$, sample size: 100, number of replications: 5,000, Model: $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$, with $\epsilon_t \overset{iid}{\sim} t(5)$. Let $\alpha = a_{12}/(a_{12}+a_{21})$ and $(1-\alpha,\alpha)$ be the stationary distribution of the hidden Markov Chain $(S_k)_k$.

a_{12}	a_{21}	α	ζ	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	MQLRT
0.4	0.6	0.4	0.6	18.6	18.9	19.0	19.2
0.2	0.8	0.2	0.8	19.2	19.7	19.9	20.1
0.3	0.7	0.3	0.6	16.6	17.0	17.2	17.4
0.2	0.8	0.2	1.4	70.5	72.9	73.5	74.6
0.1	0.9	0.1	2.0	80.7	81.9	82.1	82.0

Finally, we compare the power of the EM-test as well as of the MQLRT for testing for homogeneity in linear switching autoregressive models with possibly switching scale parameter under the alternative.

DGP 4: $X_t = 0.5X_{t-1} + (\mathbb{1}_{\{S_t=1\}} + 2\cdot\mathbb{1}_{\{S_t=2\}})\epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$.

Model 4: $X_t = \zeta + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$.

The results are shown in Table 2.13. The MQLRT and the EM-test have comparable power. Here, the power of our tests only depends on the stationary distribution of the hidden Markov chain and not on the particular transition probabilities. In contrast, if the switching mechanism is incorporated into the autoregressive part of the model (see e.g. Table 2.10 or Table 3.3) the tests have highest power when the Markov chain reduces to an i.i.d. sample, i.e. $a_{12} = 1 - a_{21}$, for strongly dependent regime, the power is smaller.

Table 2.13: Nominal level: 5%; DGP: $X_t = 0.5X_{t-1} + (\mathbb{1}_{\{S_t=1\}} + 2\cdot\mathbb{1}_{\{S_t=2\}})\epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, sample size: 200, number of replications: 5,000, Model: $X_t = \zeta + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, with $\epsilon_t \overset{iid}{\sim} N(0,1)$. Let $\alpha = a_{12}/(a_{12}+a_{21})$ and $(1-\alpha,\alpha)$ be the stationary distribution of the hidden Markov Chain $(S_k)_k$.

a_{12}	a_{21}	α	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	MQLRT
0.1	0.1	0.5	77.1	77.0	76.8	79.6
0.3	0.3		78.6	78.3	78.1	81.6
0.5	0.5		80.0	79.9	79.7	82.9
0.7	0.7		79.6	79.5	79.4	82.7
0.9	0.9		80.7	80.4	80.3	82.8
0.4	0.6	0.4	83.8	83.4	83.2	86.8
0.2	0.3		83.9	83.7	83.7	86.1
0.2	0.8	0.2	79.1	78.8	78.8	80.0
0.15	0.6		79.9	79.7	79.7	81.6
0.1	0.4		77.2	77.0	77.0	79.2

2.5 Proofs

To prove Theorem 2.1, we need the following lemma.

Lemma 2.9. *Given Assumption 2.2 and* $\delta > 0$*, let* $\alpha \in [\delta, 1-\delta]$*. Then*

$$E \log \left(\frac{g_{mix}(X_1|X_0^p; \boldsymbol{\psi})}{g(X_1|X_0^p; \vartheta_0, \boldsymbol{\eta}_0)} \right) \leq 0$$

with equality if and only if $g(x_1|x_0^p; \vartheta_i, \boldsymbol{\eta}) = g(x_1|x_0^p; \vartheta_0, \boldsymbol{\eta}_0)$ *Leb. – a.s.,* $i = 1, 2$.

Proof. Using Jensen's inequality and Assumption 2.2 we get

$$\begin{aligned} & E \log \left(\frac{g_{\text{mix}}(X_1|X_0^p; \boldsymbol{\psi})}{g(X_1|X_0^p; \vartheta_0, \boldsymbol{\eta}_0)} \right) \\ \leq \; & \log E \left(\frac{g_{\text{mix}}(X_1|X_0^p; \boldsymbol{\psi})}{g(X_1|X_0^p; \vartheta_0, \boldsymbol{\eta}_0)} \right) \\ \leq \; & 0 \end{aligned}$$

with equality if and only if $g(x_1|x_0^p; \vartheta_i, \boldsymbol{\eta}) = g(x_1|x_0^p; \vartheta_0, \boldsymbol{\eta}_0)$ Leb. $-$ $a.s.$, $i = 1, 2$. □

Proof of Theorem 2.1. (i) Since $(X_t)_t$ is stationary and ergodic, $(g(X_t|X_{t-1}^p; \vartheta, \boldsymbol{\phi}))_t$ is stationary and ergodic (cf. Krengel 1985, Prop. 1.4.3). By Assumption 2.3 and the ergodic theorem,

$$\frac{1}{n}\{l_n(1/2, \vartheta, \vartheta, \boldsymbol{\eta})\} \rightarrow E \log(g(X_1|X_0^p; \vartheta, \boldsymbol{\eta}))$$

holds almost surely for every fixed $(\vartheta, \boldsymbol{\eta}) \in \Theta \times \boldsymbol{H}$. As in Ferguson (1996), one can show that

$$\frac{1}{n}\{l_n(1/2, \vartheta, \vartheta, \boldsymbol{\eta})\} \rightarrow E \log(g(X_1|X_0^p; \vartheta, \boldsymbol{\eta})) \tag{2.5.1}$$

almost surely and uniformly over $(\vartheta, \boldsymbol{\eta}) \in \Theta \times \boldsymbol{H}$. The claim follows by Theorem 1 in Frydman (1980) using Lemma 2.9 and Assumption 2.2.

(ii) Let

$$Q(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) = E \log \left(\frac{g_{\text{mix}}(X_1|X_0^p; \boldsymbol{\psi})}{g(X_1|X_0^p; \vartheta_0, \boldsymbol{\eta}_0)} \right).$$

From Assumption 2.5 we have $R_n = O_P(1)$. Using $0 \leq M_n \leq R_n$ and the properties of the penalty function $p(\alpha)$ we get $0 \leq M_n - 2\{p(\widehat{\alpha}^*) - p(1/2)\} \leq R_n$ and therefore $p(\widehat{\alpha}^*) = O_P(1)$. Therefore there exists an $\delta > 0$ for which $P(\delta \leq \widehat{\alpha}^* \leq 1-\delta) \rightarrow 1$, $n \rightarrow \infty$, holds and we can suppose that $\alpha \in [\delta, 1-\delta]$. By the ergodic theorem and Assumption 2.4 we get under the null distribution

$$\frac{1}{n}\{pl_n(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) - pl_n(1/2, \vartheta_0, \vartheta_0, \boldsymbol{\eta}_0)\} \rightarrow Q(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) \tag{2.5.2}$$

almost surely and uniformly over $(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) \in [\delta, 1-\delta] \times \Theta^2 \times \boldsymbol{H}$. Let ω be a point in the sample space for which (2.5.2) is true and note that the set of all such points has probability 1.

Suppose for a ω the claim of the theorem is not true and, for example (the procedure for the other parameters is the same), $\widehat{\vartheta}_1^*$ does not converge to ϑ_0. There must exist a subsequence (n') such that $\widehat{\vartheta}_{1n'}^* \to \vartheta' \neq \vartheta_0$. Consider

$$\Omega' = \{(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) : |\vartheta_1 - \vartheta_0| \geq \epsilon, \alpha \in [\delta, 1-\delta]\},$$

where $\epsilon = |\vartheta' - \vartheta_0|/2$. Then for all large n', $(\widehat{\alpha}^*, \widehat{\vartheta}_1^*, \widehat{\vartheta}_2^*, \widehat{\boldsymbol{\eta}}^*)$ at the sample point ω, belongs to the subset. By Assumption 2.2 and Lemma 2.9 $Q(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) < 0$ for all $(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) \in \Omega'$. It then follows that

$$pl_{n'}(\widehat{\alpha}^*, \widehat{\vartheta}_1^*, \widehat{\vartheta}_2^*, \widehat{\boldsymbol{\eta}}^*) - pl_{n'}(1/2, \widehat{\vartheta}_0, \widehat{\vartheta}_0, \widehat{\boldsymbol{\eta}}_0) < 0$$

for all large enough n'. But this is a contradiction to $(\widehat{\alpha}^*, \widehat{\vartheta}_1^*, \widehat{\vartheta}_2^*, \widehat{\boldsymbol{\eta}}^*)$ being modified maximum likelihood estimator, and so $\widehat{\vartheta}_{1n'}^* \to \vartheta_0$ on ω. Thus $\widehat{\vartheta}_{1n'}^* \to \vartheta_0$ almost surely. □

Proof of Theorem 2.2. Let

$$\begin{aligned}
r_n(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) &= 2\{pl_n(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) - pl_n(1/2, \widehat{\vartheta}_0, \widehat{\vartheta}_0, \widehat{\boldsymbol{\eta}}_0)\}, \\
r_{1n}(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) &= 2\{pl_n(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) - pl_n(1/2, \vartheta_0, \vartheta_0, \boldsymbol{\eta}_0)\}, \\
r_{2n} &= 2\{pl_n(1/2, \vartheta_0, \vartheta_0, \boldsymbol{\eta}_0) - pl_n(1/2, \widehat{\vartheta}_0, \widehat{\vartheta}_0, \widehat{\boldsymbol{\eta}}_0)\}.
\end{aligned}$$

Therefore, $M_n = r_n(\widehat{\alpha}^*, \widehat{\vartheta}_1^*, \widehat{\vartheta}_2^*, \widehat{\boldsymbol{\eta}}^*)$ and $r_n(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) = r_{1n}(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) + r_{2n}$. We first examine r_{1n} : Expand

$$r_{1n}(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) = 2\sum_{i=1}^{n} \log(1+\delta_i) + 2p(\alpha) - 2p(1/2) \tag{2.5.3}$$

with

$$\delta_i = (1-\alpha)\left\{\frac{g(X_i|X_{i-1}^p; \vartheta_1, \boldsymbol{\eta})}{g(X_i|X_{i-1}^p; \vartheta_0, \boldsymbol{\eta}_0)} - 1\right\} + \alpha\left\{\frac{g(X_i|X_{i-1}^p; \vartheta_2, \boldsymbol{\eta})}{g(X_i|X_{i-1}^p; \vartheta_0, \boldsymbol{\eta}_0)} - 1\right\}. \tag{2.5.4}$$

We can write δ_i as

$$\begin{aligned}
\delta_i \;&=\; (1-\alpha)(\vartheta_1 - \vartheta_0)\frac{g(X_i|X_{i-1}^p; \vartheta_1, \boldsymbol{\eta}) - g(X_i|X_{i-1}^p; \vartheta_0, \boldsymbol{\eta})}{(\vartheta_1 - \vartheta_0)g(X_i|X_{i-1}^p; \vartheta_0, \boldsymbol{\eta}_0)} \\
&\quad + \alpha(\vartheta_2 - \vartheta_0)\frac{g(X_i|X_{i-1}^p; \vartheta_2, \boldsymbol{\eta}) - g(X_i|X_{i-1}^p; \vartheta_0, \boldsymbol{\eta})}{(\vartheta_2 - \vartheta_0)g(X_i|X_{i-1}^p; \vartheta_0, \boldsymbol{\eta}_0)}
\end{aligned}$$

$$
\begin{aligned}
&+(\eta_1-\eta_{1,0})\frac{g(X_i|X_{i-1}^p;\vartheta_0,\boldsymbol{\eta})-g(X_i|X_{i-1}^p;\vartheta_0,\eta_{1,0},\eta_2,\ldots,\eta_d)}{(\eta_1-\eta_{1,0})g(X_i|X_{i-1}^p;\vartheta_0,\boldsymbol{\eta}_0)}\\
&+(\eta_2-\eta_{2,0})\frac{g(X_i|X_{i-1}^p;\vartheta_0,\eta_{1,0},\eta_2,\ldots,\eta_d)-g(X_i|X_{i-1}^p;\vartheta_0,\eta_{1,0},\eta_{2,0},\eta_3,\ldots,\eta_d)}{(\eta_2-\eta_{2,0})g(X_i|X_{i-1}^p;\vartheta_0,\boldsymbol{\eta}_0)}\\
&\vdots\\
&+(\eta_d-\eta_{d,0})\frac{g(X_i|X_{i-1}^p;\vartheta_0,\eta_{1,0},\ldots,\eta_{d-1,0},\eta_d)-g(X_i|X_{i-1}^p;\vartheta_0,\boldsymbol{\eta}_0)}{(\eta_d-\eta_{d,0})g(X_i|X_{i-1}^p;\vartheta_0,\boldsymbol{\eta}_0)}\\
= \;&(1-\alpha)(\vartheta_1-\vartheta_0)Y_i(\vartheta_1,\boldsymbol{\eta})+\alpha(\vartheta_2-\vartheta_0)Y_i(\vartheta_2,\boldsymbol{\eta})+\\
&+(\eta_1-\eta_{1,0})U_i^{\eta_1}(\boldsymbol{\eta})+\ldots+(\eta_d-\eta_{d,0})U_i^{\eta_d}(\eta_{1,0},\ldots,\eta_{d-1,0},\eta_d), \qquad (2.5.5)
\end{aligned}
$$

where $U_i^{\eta_j}(\cdot)$ is defined in (2.2.6). Now, for $j=1,2$,

$$
\begin{aligned}
Y_i(\vartheta_j,\boldsymbol{\eta}) \;=\;& Y_i(\vartheta_j,\boldsymbol{\eta})-Y_i(\vartheta_j,\eta_1,\ldots,\eta_{d-1},\eta_{d,0})\\
&+Y_i(\vartheta_j,\eta_1,\ldots,\eta_{d-1},\eta_{d,0})-Y_i(\vartheta_j,\eta_1,\ldots,\eta_{d-2},\eta_{d-1,0},\eta_{d,0})\\
&\vdots\\
&+Y_i(\vartheta_j,\eta_1,\eta_{2,0},\ldots,\eta_{d-1,0},\eta_{d,0})-Y_i(\vartheta_j,\boldsymbol{\eta}_0)\\
&+(\vartheta_j-\vartheta_0)(Z_i(\vartheta_j)-Z_i)\\
&+(\vartheta_j-\vartheta_0)Z_i+Y_i \qquad (2.5.6)
\end{aligned}
$$

and

$$
\begin{aligned}
U_i^{\eta_d}(\eta_{1,0},\ldots,\eta_{d-1,0},\eta_d) &= U_i^{\eta_d}(\eta_{1,0},\ldots,\eta_{d-1,0},\eta_d)-U_i^{\eta_d}+U_i^{\eta_d}\\
U_i^{\eta_{d-1}}(\eta_{1,0},\ldots,\eta_{d-2,0},\eta_{d-1},\eta_d) &= U_i^{\eta_{d-1}}(\eta_{1,0},\ldots,\eta_{d-2,0},\eta_{d-1},\eta_d)\\
&\quad -U_i^{\eta_{d-1}}(\eta_{1,0},\ldots,\eta_{d-1,0},\eta_d)\\
&\quad +U_i^{\eta_{d-1}}(\eta_{1,0},\ldots,\eta_{d-1,0},\eta_d)-U_i^{\eta_{d-1}}\\
&\quad +U_i^{\eta_{d-1}}\\
&\vdots\\
U_i^{\eta_1}(\boldsymbol{\eta}) &= U_i^{\eta_1}(\boldsymbol{\eta})-U_i^{\eta_1}(\eta_1,\ldots,\eta_{d-1},\eta_{d,0})\\
&\quad +U_i^{\eta_1}(\eta_1,\ldots,\eta_{d-1},\eta_{d,0})\\
&\quad -U_i^{\eta_1}(\eta_1,\ldots,\eta_{d-2},\eta_{d-1,0},\eta_{d,0})\\
&\quad \vdots\\
&\quad +U_i^{\eta_1}(\eta_1,\eta_{2,0},\ldots,\eta_{d-1,0},\eta_{d,0})-U_i^{\eta_1}\\
&\quad +U_i^{\eta_1}. \qquad (2.5.7)
\end{aligned}
$$

Plugging (2.5.6) and (2.5.7) into (2.5.5), we can write

$$
\delta_i=(\eta_1-\eta_{1,0})U_i^{\eta_1}+\ldots+(\eta_d-\eta_{d,0})U_i^{\eta_d}+m_1Y_i+m_2Z_i+\epsilon_{in}, \qquad (2.5.8)
$$

where

$$m_1 = (1-\alpha)(\vartheta_1 - \vartheta_0) + \alpha(\vartheta_2 - \vartheta_0), \qquad m_2 = (1-\alpha)(\vartheta_1 - \vartheta_0)^2 + \alpha(\vartheta_2 - \vartheta_0)^2$$

and ϵ_{in} is a remainder term. Note at this stage that in each of the sequences the variables $(U_i^{\eta_j})_{i\geq 1}$, $j = 1, \ldots d$, $(Y_i)_{i\geq 1}$ and $(Z_i)_{i\geq 1}$ form square integrable (Assumption 2.4) stationary martingale difference sequences w.r.t. the filtration generated by the observations (X_i). Let $\epsilon_n = \sum_{i=1}^n \epsilon_{in}$. By Assumption 2.7,

$$\begin{aligned}
\epsilon_n \;=\; & \sqrt{n}(\eta_d - \eta_{d,0})^2 O_P(1) \\
& + \sqrt{n}(\eta_{d-1} - \eta_{d-1,0})\Big(\sum_{j=d-1}^{d}(\eta_j - \eta_{j,0})\Big) O_P(1) \\
& \vdots \\
& + \sqrt{n}(\eta_1 - \eta_{1,0})\Big(\sum_{j=1}^{d}(\eta_j - \eta_{j,0})\Big) O_P(1) \\
& + \sqrt{n}(1-\alpha)(\vartheta_1 - \vartheta_0)\Big(\sum_{j=1}^{d}(\eta_j - \eta_{j,0})\Big) O_P(1) \\
& + \sqrt{n}\alpha(\vartheta_2 - \vartheta_0)\Big(\sum_{j=1}^{d}(\eta_j - \eta_{j,0})\Big) O_P(1) \\
& + \sqrt{n}(1-\alpha)(\vartheta_1 - \vartheta_0)^3 O_P(1) + \sqrt{n}\alpha(\vartheta_2 - \vartheta_0)^3 O_P(1).
\end{aligned}$$

Let us now restrict our attention to a small neighborhood of $(\eta_{1,0}, \ldots, \eta_{d,0}, \vartheta_0)$ as suggested by the consistency results in Theorem 2.1(ii). Therefore, we may regard $\eta_1 - \eta_{1,0}, \ldots, \eta_d - \eta_{d,0}, \vartheta_1 - \vartheta_0, \vartheta_2 - \vartheta_0$ as $o_P(1)$ and we get

$$\begin{aligned}
\epsilon_n \;=\; & \sqrt{n}(\eta_d - \eta_{d,0}) o_P(1) + \sqrt{n}(\eta_{d-1} - \eta_{d-1,0}) o_P(1) + \ldots + \sqrt{n}(\eta_1 - \eta_{1,0}) o_P(1) \\
& + \sqrt{n}(1-\alpha)(\vartheta_1 - \vartheta_0) o_P(1) + \sqrt{n}\alpha(\vartheta_2 - \vartheta_0) o_P(1) \\
& + \sqrt{n}(1-\alpha)(\vartheta_1 - \vartheta_0)^2 o_P(1) + \sqrt{n}\alpha(\vartheta_2 - \vartheta_0)^2 o_P(1).
\end{aligned}$$

Since $|x| \leq 1 + x^2$, we obtain

$$|\epsilon_n| \leq n\{(\eta_1 - \eta_{1,0})^2 + \ldots + (\eta_d - \eta_{d,0})^2 + m_1^2 + m_2^2\} o_p(1) + o_P(1).$$

By Assumption 2.6 there is a $\lambda > 0$ such that for all $(\alpha_1, \ldots, \alpha_{d+2}) \in \mathbb{R}^{d+2} \setminus \{\mathbf{0}\}$ we have

$$E\{\alpha_1 U_1^{\eta_1} + \ldots + \alpha_d U_1^{\eta_d} + \alpha_{d+1} Y_1 + \alpha_{d+2} Z_1\}^2 \geq \lambda(\alpha_1^2 + \ldots + \alpha_{d+2}^2). \tag{2.5.9}$$

The ergodic theorem, Assumption 2.4 and (2.5.9) imply

$$\begin{aligned}
& \frac{\sum_{i=1}^n |(\eta_1-\eta_{1,0})U_i^{\eta_1} + \ldots (\eta_d-\eta_{d,0})U_i^{\eta_d} + m_1 Y_i + m_2 Z_i|^3}{\sum_{i=1}^n ((\eta_1-\eta_{1,0})U_i^{\eta_1} + \ldots (\eta_d-\eta_{d,0})U_i^{\eta_d} + m_1 Y_i + m_2 Z_i)^2} \\
= & \frac{E|\eta_1-\eta_{1,0})U_1^{\eta_1} + \ldots (\eta_d-\eta_{d,0})U_1^{\eta_d} + m_1 Y_1 + m_2 Z_1|^3}{E((\eta_1-\eta_{1,0})U_1^{\eta_1} + \ldots (\eta_d-\eta_{d,0})U_1^{\eta_d} + m_1 Y_1 + m_2 Z_1)^2} O_P(1) \\
\le & \frac{|\eta_1-\eta_{1,0}|^3 + \ldots + |\eta_d-\eta_{d,0}|^3 + |m_1|^3 + |m_2|^3}{(\eta_1-\eta_{1,0})^2 + \ldots + (\eta_d-\eta_{d,0})^2 + m_1^2 + m_2^2} O_P(1) \\
\le & \{|\eta_1-\eta_{1,0}| + \ldots + |\eta_d-\eta_{d,0}| + |m_1| + |m_2|\} O_P(1) = o_P(1).
\end{aligned}$$

Therefore, using the properties of the penalty function in the second step, we get

$$\begin{aligned}
& r_{1n}(\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}) \\
= & 2\sum_{i=1}^n \log(1+\delta_i) + 2p(\alpha) - 2p(1/2) \\
\le & 2\sum_{i=1}^n \delta_i - \sum_{i=1}^n \delta_i^2 + \frac{2}{3}\sum_{i=1}^n \delta_i^3 \\
\le & 2\sum_{i=1}^n \{(\eta_1-\eta_{1,0})U_i^{\eta_1} + \ldots + (\eta_d-\eta_{d,0})U_i^{\eta_d} + m_1 Y_i + m_2 Z_i\} \\
& -\sum_{i=1}^n \{(\eta_1-\eta_{1,0})U_i^{\eta_1} + \ldots + (\eta_d-\eta_{d,0})U_i^{\eta_d} + m_1 Y_i + m_2 Z_i\}^2 \{1 + o_P(1)\} \\
& + o_P(1).
\end{aligned}$$

We orthogonalize

$$\begin{aligned}
\tilde{U}_i^{\eta_1} &= U_i^{\eta_1}, \qquad \tilde{U}_i^{\eta_2} = U_i^{\eta_2} - \frac{\sum_{k=1}^n \tilde{U}_k^{\eta_1} U_k^{\eta_2}}{\sum_{k=1}^n \left(\tilde{U}_k^{\eta_1}\right)^2} \tilde{U}_i^{\eta_1}, \ldots \qquad (2.5.10) \\
\tilde{Y}_i &= Y_i - \sum_{j=1}^d \frac{\sum_{k=1}^n \tilde{U}_k^{\eta_j} Y_k}{\sum_{k=1}^n \left(\tilde{U}_k^{\eta_j}\right)^2} \tilde{U}_i^{\eta_j}, \qquad \tilde{Z}_i = Z_i - \frac{\sum_{k=1}^n Z_k \tilde{Y}_k}{\sum_{k=1}^n \tilde{Y}_k^2} \tilde{Y}_i - \sum_{j=1}^d \frac{\sum_{k=1}^n \tilde{U}_k^{\eta_j} Y_k}{\sum_{k=1}^n \left(\tilde{U}_k^{\eta_j}\right)^2} \tilde{U}_i^{\eta_j}.
\end{aligned}$$

Therefore

$$\begin{aligned}
& (\eta_1-\eta_{1,0})U_i^{\eta_1} + \ldots (\eta_d-\eta_{d,0})U_i^{\eta_d} + m_1 Y_i + m_2 Z_i \\
= & t_1 \tilde{U}_i^{\eta_1} + t_2 \tilde{U}_i^{\eta_2} + \ldots t_d \tilde{U}_i^{\eta_d} + t_{d+1} \tilde{Y}_i + t_{d+2} \tilde{Z}_i
\end{aligned}$$

with some coefficients t_i, where in particular $t_{d+2} = m_2$.

Computing the maximum of the quadratic function

$$\begin{aligned} q(t_1,\ldots,t_{d+2}) =& 2\sum_{i=1}^{n}\{t_1\tilde{U}_i^{\eta_1}+\ldots t_d\tilde{U}_i^{\eta_d}+t_{d+1}\tilde{Y}_i+t_{d+2}\tilde{Z}_i\} \\ &-\sum_{i=1}^{n}\{t_1\tilde{U}_i^{\eta_1}+\ldots t_d\tilde{U}_i^{\eta_d}+t_{d+1}\tilde{Y}_i+t_{d+2}\tilde{Z}_i\}^2 \end{aligned}$$

we get an asymptotic upper bound for r_{1n} as follows. Due to the constraint $t_{d+2}\geq 0$,

$$\begin{aligned} (\tilde{t}_1,\ldots,\tilde{t}_{d+2}) &= \arg\max_{t_1,\ldots,t_{d+2}} q(t_1,\ldots,t_{d+2}) \\ &= \left(\frac{\sum\tilde{U}_i^{\eta_1}}{\sum\left(\tilde{U}_i^{\eta_1}\right)^2},\ldots,\frac{\sum\tilde{U}_i^{\eta_d}}{\sum\left(\tilde{U}_i^{\eta_d}\right)^2},\frac{\sum\tilde{Y}_i}{\sum\tilde{Y}_i^2},\frac{(\sum\tilde{Z}_i)^+}{\sum\tilde{Z}_i^2}\right) \end{aligned} \tag{2.5.11}$$

and therefore an upper bound for r_{1n} is given by

$$r_{1n}(\widehat{\alpha}^*,\widehat{\vartheta}_1^*,\widehat{\vartheta}_2^*,\widehat{\boldsymbol{\eta}}^*)\leq\frac{(\sum\tilde{U}_i^{\eta_1})^2}{\sum\left(\tilde{U}_i^{\eta_1}\right)^2}+\ldots+\frac{(\sum\tilde{U}_i^{\eta_d})^2}{\sum\left(\tilde{U}_i^{\eta_d}\right)^2}+\frac{(\sum\tilde{Y}_i)^2}{\sum\tilde{Y}_i^2}+\frac{((\sum\tilde{Z}_i)^+)^2}{\sum\tilde{Z}_i^2}+o_P(1). \tag{2.5.12}$$

For $\alpha=1/2$ and the values $\tilde{\vartheta}_1^*,\tilde{\vartheta}_2^*$ and $\tilde{\boldsymbol{\eta}}^*$ given implicitly in (2.5.11) we see that this upper bound is attained.

Expanding r_{2n} in a similar way as r_{1n} (see below),

$$\begin{aligned} -r_{2n} &= 2\left\{pl_n(1/2,\widehat{\vartheta}_0,\widehat{\vartheta}_0,\widehat{\boldsymbol{\eta}}_0)-pl_n(1/2,\vartheta_0,\vartheta_0,\boldsymbol{\eta}_0)\right\} \\ &= \frac{(\sum\tilde{U}_i^{\eta_1})^2}{\sum\left(\tilde{U}_i^{\eta_1}\right)^2}+\ldots+\frac{(\sum\tilde{U}_i^{\eta_d})^2}{\sum\left(\tilde{U}_i^{\eta_d}\right)^2}+\frac{(\sum\tilde{Y}_i)^2}{\sum\tilde{Y}_i^2}+o_P(1). \end{aligned}$$

Therefore,

$$M_n=\frac{\left((\sum\tilde{Z}_i)^+\right)^2}{\sum\tilde{Z}_i^2}+o_P(1).$$

Let $(\hat{U}_i^{\eta_j}),(\hat{Y}_i),(\hat{Z}_i)$ be the square integrable stationary martingale difference sequences obtained by replacing in (2.5.10) at each stage the empirical scalar products by their expected versions (e.g. $\hat{U}_i^{\eta_2}=U_i^{\eta_2}-\frac{EU_1^{\eta_1}U_1^{\eta_2}}{\left(EU_1^{\eta_1}\right)^2}U_i^{\eta_1}$). Then $n^{-1/2}\sum_i(\tilde{Z}_i-\hat{Z}_i)=o_P(1)$ and therefore our result follows from the ergodic theorem (applied to the denominator) and the central limit theorem for stationary ergodic martingale difference sequences (applied to the numerator). □

Expansion of r_{2n}. We expand

$$r_{2n} = 2\sum_{i=1}^{n} \log(1+\delta_i), \qquad \delta_i = \frac{g(X_i|X_{i-1}^p;\vartheta,\boldsymbol{\eta})}{g(X_i|X_{i-1}^p;\vartheta_0,\boldsymbol{\eta}_0)} - 1. \tag{2.5.13}$$

Write δ_i as

$$\delta_i \;=\; (\vartheta-\vartheta_0)Y_i(\vartheta,\boldsymbol{\eta}) + (\eta_1-\eta_{1,0})U_i^{\eta_1}(\boldsymbol{\eta}) + \ldots + (\eta_d-\eta_{d,0})U_i^{\eta_d}(\eta_{1,0},\ldots,\eta_{d-1,0},\eta_d).$$

Now,

$$\begin{aligned}
Y_i(\vartheta,\boldsymbol{\eta}) \;=\;& \Big(Y_i(\vartheta,\boldsymbol{\eta}) - Y_i(\vartheta,\eta_1,\ldots,\eta_{d-1},\eta_{d,0})\Big)\\
&+\Big(Y_i(\vartheta,\eta_1,\ldots,\eta_{d-1},\eta_{d,0}) - Y_i(\vartheta,\eta_1,\ldots,\eta_{d-2},\eta_{d-1,0},\eta_{d,0})\Big)\\
&\ldots\\
&+\Big(Y_i(\vartheta,\eta_1,\eta_{2,0},\ldots,\eta_{d-1,0},\eta_{d,0}) - Y_i(\vartheta,\boldsymbol{\eta}_0)\Big)\\
&+\Big(Y_i(\vartheta,\boldsymbol{\eta}_0) - Y_i\Big) + Y_i
\end{aligned}$$

and

$$\begin{aligned}
U_i^{\eta_d}(\eta_{1,0},\ldots,\eta_{d-1,0},\eta_d) \;=\;& U_i^{\eta_d}(\eta_{1,0},\ldots,\eta_{d-1,0},\eta_d) - U_i^{\eta_d} + U_i^{\eta_d}\\
U_i^{\eta_{d-1}}(\eta_{1,0},\ldots,\eta_{d-2,0},\eta_{d-1},\eta_d) \;=\;& \big(U_i^{\eta_{d-1}}(\eta_{1,0},\ldots,\eta_{d-2,0},\eta_{d-1},\eta_d)\\
&-U_i^{\eta_{d-1}}(\eta_{1,0},\ldots,\eta_{d-1,0},\eta_d)\big)\\
&+U_i^{\eta_{d-1}}(\eta_{1,0},\ldots,\eta_{d-1,0},\eta_d) - U_i^{\eta_{d-1}}\\
&+U_i^{\eta_{d-1}}\\
&\vdots\\
U_i^{\eta_1}(\boldsymbol{\eta}) \;=\;& U_i^{\eta_1}(\boldsymbol{\eta}) - U_i^{\eta_1}(\eta_1,\ldots,\eta_{d-1},\eta_{d,0})\\
&+\big(U_i^{\eta_1}(\eta_1,\ldots,\eta_{d-1},\eta_{d,0})\\
&-U_i^{\eta_1}(\eta_1,\ldots,\eta_{d-2},\eta_{d-1,0},\eta_{d,0})\big)\\
&\vdots\\
&+U_i^{\eta_1}(\eta_1,\eta_{2,0},\ldots,\eta_{d-1,0},\eta_{d,0}) - U_i^{\eta_1}\\
&+U_i^{\eta_1}.
\end{aligned}$$

Then

$$\delta_i = (\eta_1-\eta_{1,0})U_i^{\eta_1} + \ldots + (\eta_d-\eta_{d,0})U_i^{\eta_d} + (\vartheta-\vartheta_0)Y_i + \epsilon_{in}, \tag{2.5.14}$$

where ϵ_{in} is a remainder term. Let $\epsilon_n = \sum_{i=1}^n \epsilon_{in}$. By Assumption 2.7 we show

$$\begin{aligned}\epsilon_n &= \sqrt{n}(\eta_d - \eta_{d,0})^2 O_P(1)\\ &+\sqrt{n}(\eta_{d-1} - \eta_{d-1,0})\{\sum_{j=d-1}^{d}(\eta_j - \eta_{j,0})\}O_P(1) + \ldots\\ &+\sqrt{n}(\eta_1 - \eta_{1,0})\{\sum_{j=1}^{d}(\eta_j - \eta_{j,0})\}O_P(1)\\ &+\sqrt{n}(\vartheta - \vartheta_0)\{\sum_{j=1}^{d}(\eta_j - \eta_{j,0})\}O_P(1).\end{aligned}$$

Let us now restrict our attention to a small neighorhood of $(\eta_{1,0}, \ldots, \eta_{d,0})$ as suggested by the consistency results in Theorem 2.1(i). Therefore, we may regard $\eta_1 - \eta_{1,0}, \ldots, \eta_d - \eta_{d,0}$ as $o_P(1)$ and we get

$$\begin{aligned}\epsilon_n &= \sqrt{n}(\eta_d - \eta_{d,0})o_P(1) + \sqrt{n}(\eta_{d-1} - \eta_{d-1,0})o_P(1) + \ldots\\ &+\sqrt{n}(\eta_1 - \eta_{1,0})o_P(1) + \sqrt{n}(\vartheta - \vartheta_0)o_P(1).\end{aligned}$$

Using $|x| \leq 1 + x^2$ we obtain

$$|\epsilon_n| \leq n\{(\eta_1 - \eta_{1,0})^2 + \ldots + (\eta_d - \eta_{d,0})^2 + (\vartheta - \vartheta_0)^2\}o_p(1) + o_P(1).$$

By Assumption 2.6, there exists a $\lambda > 0$

$$E\{\alpha_1 U_1^{\eta_1} + \ldots \alpha_d U_1^{\eta_d} + \alpha_{d+1} Y_1\}^2 \geq \lambda(\alpha_1^2 + \ldots + \alpha_{d+1}^2) \tag{2.5.15}$$

for all $(\alpha_1, \ldots, \alpha_{d+1}) \in \mathbb{R}^{d+1} \setminus \{\mathbf{0}\}$. The ergodic theorem, Assumption 2.4 and (2.5.9) imply

$$\begin{aligned}&\frac{\sum_{i=1}^n |(\eta_1 - \eta_{1,0})U_i^{\eta_1} + \ldots (\eta_d - \eta_{d,0})U_i^{\eta_d} + (\vartheta - \vartheta_0)Y_i|^3}{\sum_{i=1}^n ((\eta_1 - \eta_{1,0})U_i^{\eta_1} + \ldots (\eta_d - \eta_{d,0})U_i^{\eta_d} + (\vartheta - \vartheta_0)Y_i)^2}\\ =\; &\frac{E|(\eta_1 - \eta_{1,0})U_1^{\eta_1} + \ldots (\eta_d - \eta_{d,0})U_1^{\eta_d} + (\vartheta - \vartheta_0)Y_1|^3}{E((\eta_1 - \eta_{1,0})U_1^{\eta_1} + \ldots (\eta_d - \eta_{d,0})U_1^{\eta_d} + (\vartheta - \vartheta_0)Y_1)^2}O_P(1)\\ \leq\; &\frac{|\eta_1 - \eta_{1,0}|^3 + \ldots + |\eta_d - \eta_{d,0}|^3 + |\vartheta - \vartheta_0|^3}{(\eta_1 - \eta_{1,0})^2 + \ldots + (\eta_d - \eta_{d,0})^2 + (\vartheta - \vartheta_0)^2}O_P(1)\\ \leq\; &\{|\eta_1 - \eta_{1,0}| + \ldots + |\eta_d - \eta_{d,0}| + |\vartheta - \vartheta_0|\}O_P(1) = o_P(1).\end{aligned}$$

Therefore, we get

$$r_{2n} = 2\sum_{i=1}^n \log(1 + \delta_i)$$

$$
\begin{aligned}
&\leq \; 2\sum_{i=1}^{n}\delta_i - \sum_{i=1}^{n}\delta_i^2 + \frac{2}{3}\sum_{i=1}^{n}\delta_i^3 \\
&\leq \; 2\sum_{i=1}^{n}\{(\eta_1-\eta_{1,0})U_i^{\eta_1} + \ldots + (\eta_d-\eta_{d,0})U_i^{\eta_d} + (\vartheta-\vartheta_0)Y_i\} \\
&\quad -\sum_{i=1}^{n}\{(\eta_1-\eta_{1,0})U_i^{\eta_1} + \ldots + (\eta_d-\eta_{d,0})U_i^{\eta_d} + (\vartheta-\vartheta_0)Y_i\}^2\{1+o_P(1)\} \\
&\quad +o_P(1).
\end{aligned}
$$

We proceed as in (2.5.10) and obtain

$$
\begin{aligned}
& (\eta_1-\eta_{1,0})U_i^{\eta_1} + \ldots (\eta_d-\eta_{d,0})U_i^{\eta_d} + (\vartheta-\vartheta_0)Y_i \\
= \; & t_1\tilde{U}_i^{\eta_1} + t_2\tilde{U}_i^{\eta_2} + \ldots t_d\tilde{U}_i^{\eta_d} + t_{d+1}\tilde{Y}_i
\end{aligned}
$$

with some coefficients t_i. Computing the maximum of the quadratic function

$$
\begin{aligned}
q(t_1,\ldots,t_{d+1}) =& 2\sum_{i=1}^{n}\{t_1\tilde{U}_i^{\eta_1} + \ldots t_d\tilde{U}_i^{\eta_d} + t_{d+1}\tilde{Y}_i\} \\
& -\sum_{i=1}^{n}\{t_1\tilde{U}_i^{\eta_1} + \ldots t_d\tilde{U}_i^{\eta_d} + t_{d+1}\tilde{Y}_i\}^2
\end{aligned}
$$

we get

$$
\begin{aligned}
(\tilde{t}_1,\ldots,\tilde{t}_{d+1}) =& \arg\max_{t_1,\ldots,t_{d+2}} q(t_1,\ldots,t_{d+1}) \\
=& \left(\frac{\sum\tilde{U}_i^{\eta_1}}{\sum\left(\tilde{U}_i^{\eta_1}\right)^2},\ldots,\frac{\sum\tilde{U}_i^{\eta_d}}{\sum\left(\tilde{U}_i^{\eta_d}\right)^2},\frac{\sum\tilde{Y}_i}{\sum\tilde{Y}_i^2}\right).
\end{aligned}
\tag{2.5.16}
$$

An upper bound for r_{2n} is then given by

$$
r_{2n} \leq \frac{(\sum\tilde{U}_i^{\eta_1})^2}{\sum\left(\tilde{U}_i^{\eta_1}\right)^2} + \ldots + \frac{(\sum\tilde{U}_i^{\eta_d})^2}{\sum\left(\tilde{U}_i^{\eta_d}\right)^2} + \frac{(\sum\tilde{Y}_i)^2}{\sum\tilde{Y}_i^2} + o_P(1).
$$

For the values $\tilde{\vartheta}$ and $\tilde{\boldsymbol{\eta}}$ which are implicitly given in (2.5.16), that upper bound is attained.

Proof of Lemma 2.2. First, we consider model (2.2.1). Let $\mu(\zeta,\phi_1,\ldots,\phi_p;x_0^p) = \zeta+\phi_1x_0+\ldots+\phi_px_{1-p}$. Then

$$
U_1^{\phi_j} \;=\; \frac{\frac{\partial}{\partial\mu}f\big(X_1;\mu(\zeta,\phi_1,\ldots,\phi_p;X_0^p),\sigma\big)X_{1-j}}{f\big(X_1;\mu(\zeta,\phi_1,\ldots,\phi_p;X_0^p),\sigma\big)}, \quad j=1,\ldots,p,
$$

$$
\begin{aligned}
U_1^{\sigma} &= \frac{\frac{\partial}{\partial\sigma} f\big(X_1; \mu(\zeta, \phi_1, \dots, \phi_p; X_0^p), \sigma\big)}{f\big(X_1; \mu(\zeta, \phi_1, \dots, \phi_p; X_0^p), \sigma\big)}, \\
Y_1 &= \frac{\frac{\partial}{\partial\mu} f\big(X_1; \mu(\zeta, \phi_1, \dots, \phi_p; X_0^p), \sigma\big)}{f\big(X_1; \mu(\zeta, \phi_1, \dots, \phi_p; X_0^p), \sigma\big)}, \\
Z_1 &= \frac{\frac{\partial^2}{\partial^2\mu} f\big(X_1; \mu(\zeta, \phi_1, \dots, \phi_p; X_0^p), \sigma\big)}{f\big(X_1; \mu(\zeta, \phi_1, \dots, \phi_p; X_0^p), \sigma\big)}.
\end{aligned}
$$

The covariance is non-degenerate if and only if these random variables are linearly independent (in L_2). Therefore, suppose that for constants b_j,

$$
b_1 Z_1 + b_2 Y_1 + b_3 U_1^{\sigma} + \sum_{j=1}^{p} b_{j+3} U_1^{\phi_j} = 0 \quad a.s. \tag{2.5.17}
$$

Since the distribution of $X_1, \dots, X_{1-p}$ is equivalent to Lebesgue measure on $\mathbb{R}^{p+1}$, in the sense that the associated probability measure and the Lebesgue measure on $\mathbb{R}^{p+1}$ are mutually absolutely continuous, (2.5.17) is equivalent to

$$
b_1 \frac{\partial^2}{\partial^2\mu} f + b_2 \frac{\partial}{\partial\mu} f + b_3 \frac{\partial}{\partial\sigma} f + \sum_{j=1}^{p} b_{j+3} \frac{\partial}{\partial\mu} f\, x_{1-j} = 0 \quad \text{Leb.} - a.s.
$$

where $f = f\big(x_1; \mu(\zeta, \phi_1, \dots, \phi_d; x_0^p), \sigma\big)$. From (2.3.3), it follows that $b_1 = b_3 = 0$ and that

$$
b_2 + \sum_{j=1}^{p} b_{j+3} x_{1-j} = 0 \quad \text{Leb.} - a.s.,
$$

so that $b_2 = b_4 = \ldots = b_{p+3} = 0$.

Let j_0 be the index of the autoregressive parameter which switches according to the hidden state S_t in model (2.2.2) and $\mu(\zeta, \phi_1, \dots, \phi_p; x_0^p) = \zeta + \phi_1 x_0 + \ldots + \phi_p x_{1-p}$. Then

$$
\begin{aligned}
U_1^{\zeta} &= \frac{\frac{\partial}{\partial\mu} f(X_1; \mu(\zeta, \phi_1, \dots, \phi_p; X_0^p), \sigma)}{f(X_1; \mu(\zeta, \phi_1, \dots, \phi_p; X_0^p), \sigma)}, \\
U_1^{\phi_\tau} &= \frac{\frac{\partial}{\partial\mu} f(X_1; \mu(\zeta, \phi_1, \dots, \phi_p; X_0^p), \sigma) X_{1-\tau}}{f(X_1; \mu(\zeta, \phi_1, \dots, \phi_p; X_0^p), \sigma)} = U_1^{\zeta} X_{1-\tau}, \quad 1 \le \tau \ne\ j_0 \le p, \\
Y_1 &= \frac{\frac{\partial}{\partial\mu} f(X_1; \mu(\zeta, \phi_1, \dots, \phi_p; X_0^p), \sigma) X_{1-j_0}}{f(X_1; \mu(\zeta, \phi_1, \dots, \phi_p; X_0^p), \sigma)} = U_1^{\zeta} X_{1-j_0}, \\
Z_1 &= \frac{\frac{\partial^2}{\partial^2\mu} f(X_1; \mu(\zeta, \phi_1, \dots, \phi_p; X_0^p), \sigma) X_{1-j_0}^2}{f(X_1; \mu(\zeta, \phi_1, \dots, \phi_p; X_0^p), \sigma)},
\end{aligned}
$$

$$U_1^{\sigma} \quad = \quad \frac{\frac{\partial}{\partial \sigma} f(X_1; \mu(\zeta, \phi_1, \ldots, \phi_p; X_0^p), \sigma)}{f(X_1; \mu(\zeta, \phi_1, \ldots, \phi_p; X_0^p), \sigma)}.$$

The covariance matrix of $(U_1^{\zeta}, U_1^{\phi_1}, \ldots, U_1^{\phi_{j_0-1}}, U_1^{\phi_{j_0+1}}, \ldots, U_1^{\phi_p}, U_1^{\sigma}, Y_1, Z_1)$ is non-degenerate if and only if these random variables are linearly independent (in L_2). Therefore, suppose that for some constants b_j

$$b_1 U_1^{\zeta} + b_2 U_1^{\sigma} + b_3 Z_1 + \sum_{\tau \neq j_0} b_{3+\tau} U_1^{\phi_\tau} + b_{3+j_0} Y_1 = 0 \quad a.s. \tag{2.5.18}$$

holds. Since the distribution of $X_1, \ldots, X_{1-p}$ is equivalent to Lebesgue measure on $\mathbb{R}^{p+1}$, (2.5.18) is equivalent to

$$b_1 \frac{\partial}{\partial \mu} f + b_2 \frac{\partial}{\partial \sigma} f + b_3 \frac{\partial^2}{\partial^2 \mu} f x_{1-j_0}^2 + \sum_{\tau=1}^{p} b_{\tau+3} \frac{\partial}{\partial \mu} f x_{1-\tau} = 0 \quad \text{Leb.} - a.s. \tag{2.5.19}$$

From equation (2.3.3), it follows that

$$\begin{aligned} b_2 &= 0, \\ b_3 x_{1-j_0}^2 &= 0 \text{ Leb.} - a.s., \\ b_1 + \sum_{\tau=1}^{p} b_{\tau+3} x_{1-\tau} &= 0 \text{ Leb.} - a.s., \end{aligned}$$

so that $b_3 = 0$ and $b_1 = b_4 = \ldots = b_{p+3} = 0$. □

Proof of Lemma 2.3. The characteristic function of the t-distribution is given by (cf. Hurst 1995)

$$\varphi(t) = \frac{K_{\frac{1}{2}\nu}\left(\sqrt{\nu}|t|\right)\left(\sqrt{\nu}|t|\right)^{\frac{1}{2}\nu}}{\Gamma\left(\frac{1}{2}\nu\right) 2^{\frac{1}{2}\nu-1}},$$

where $\Gamma(\cdot)$ is the Gamma function and $K_p(\cdot)$ is the modified Bessel function of the second kind and order p (cf. Andrews 1986, Chp. 6). Therefore, the characteristic function of the corresponding location-scale family is

$$\varphi(t; \mu, \sigma) = e^{i\mu t}\, \varphi(\sigma t) = e^{i\mu t} \frac{K_m\left(\sqrt{\nu}\sigma|t|\right)\left(\sqrt{\nu}\sigma|t|\right)^m}{\Gamma(m)\, 2^{m-1}}, \tag{2.5.20}$$

where we put $m = \frac{1}{2}\nu$. The partial derivatives are

$$\frac{\partial \varphi(t; \mu, \sigma)}{\partial \mu} \quad = \quad it\, e^{i\mu t}\, \frac{K_m\left(\sqrt{\nu}\sigma|t|\right)\left(\sqrt{\nu}\sigma|t|\right)^m}{\Gamma(m)\, 2^{m-1}},$$

$$\frac{\partial^2\varphi(t;\mu,\sigma)}{\partial^2\mu} = -t^2\, e^{i\mu t}\, \frac{K_m\left(\sqrt{\nu}\sigma|t|\right)\left(\sqrt{\nu}\sigma|t|\right)^m}{\Gamma(m)\,2^{m-1}}$$

and

$$\begin{aligned}\frac{\partial\varphi(t;\mu,\sigma)}{\partial\sigma} &= -|t|\; e^{i\mu t}\frac{K_{m-1}(\sqrt{\nu}\sigma|t|)\sqrt{\nu}\,(\sqrt{\nu}\sigma|t|)^m}{\Gamma(m)2^{m-1}},\\ \frac{\partial^2\varphi(t;\mu,\sigma)}{\partial^2\sigma} &= \frac{|t|\,\sqrt{\nu}e^{i\mu t}}{\Gamma(m)2^{m-1}}\left(\sqrt{\nu}|t|\right)^m\sigma^{m-1}\Big(\sqrt{\nu}\sigma|t|K_{m-2}(\sqrt{\nu}\sigma|t|)-K_{m-1}(\sqrt{\nu}\sigma|t|)\Big),\end{aligned}$$

cf. Andrews (1986).

(i). Taking the Fourier transform in (2.3.4) and interchanging integral and derivative gives

$$a_1\frac{\partial\varphi(t;\mu,\sigma)}{\partial\mu}+a_2\frac{\partial^2\varphi(t;\mu,\sigma)}{\partial^2\mu}+a_3\frac{\partial\varphi(t;\mu,\sigma)}{\partial\sigma}=0 \quad \text{for all } t\in\mathbb{R}. \tag{2.5.21}$$

Plugging the partial derivatives into (2.5.21), dividing by $te^{i\mu t}(\sqrt{\nu}|t|)^m\,\sigma^{m-1}/\big(\Gamma(m)2^{m-1}\big)$, and putting $x=\sqrt{\nu}\sigma|t|$ gives

$$a_1\, i\sigma\; K_m(x) - a_2\sigma t\; K_m(x) - a_3\sqrt{\nu}\sigma\,\mathrm{sign}(t)K_{m-1}(x) = 0, \quad t\in\mathbb{R}. \tag{2.5.22}$$

Choosing $t=1$ and $t=-1$ and adding, we get $a_1=0$. Next, dividing by $t\,K_m(x)$ and letting $t\to\infty$ (hence $x\to\infty$), since $K_{m-1}(x)/K_m(x)\to 1$ (Andrews 1986), we get $a_2=0$ and finally $a_3=0$.

(ii). Taking the Fourier transform in (2.3.5) and interchanging integral and derivative gives

$$b_1\frac{\partial\varphi(t;\mu,\sigma)}{\partial\mu}+b_2\frac{\partial\varphi(t;\mu,\sigma)}{\partial\sigma}+b_3\frac{\partial^2\varphi(t;\mu,\sigma)}{\partial^2\sigma}=0 \quad \text{for all } t\in\mathbb{R}. \tag{2.5.23}$$

Plugging the partial derivatives into (2.5.23) and dividing by $te^{i\mu t}(\sqrt{\nu}|t|)^m\,\sigma^{m-1}/\big(\Gamma(m)2^{m-1}\big)$, and putting $x=\sqrt{\nu}\sigma|t|$ gives

$$b_1\, i\sigma\; K_m(x) - b_2\sqrt{\nu}\sigma\,\mathrm{sign}(t)K_{m-1}(x) + b_3\sqrt{\nu}\,\mathrm{sign}(t)\big(xK_{m-2}(x)-K_{m-1}(x)\big) = 0, \quad t\in\mathbb{R}. \tag{2.5.24}$$

Choosing $t=1$ and $t=-1$ and adding, we get $b_1=0$. Therefore equation (2.5.24) reduces to

$$b_2\sigma K_{m-1}(x) - b_3\big(xK_{m-2}(x)-K_{m-1}(x)\big) = 0, \quad t\in\mathbb{R}. \tag{2.5.25}$$

Dividing by $x\,K_{m-2}(x)$ and letting $x\to\infty$, since $K_{m-1}(x)/K_{m-2}(x)\to 1$ (Andrews 1986), we get $b_3=0$ and therefore $b_2=0$.

□

Proof of Lemma 2.4. The proof basically follows the proof of Lemma 2.2 for model (2.2.2) up to equation (2.5.19). For the normal distribution (2.5.19) is equivalent to

$$b_1 \frac{\partial}{\partial \mu} f + b_2 \frac{\partial}{\partial \sigma} f + \frac{b_3}{\sigma} \frac{\partial}{\partial \sigma} f x_{1-j_0}^2 + \sum_{\tau=1}^{p} b_{\tau+3} \frac{\partial}{\partial \mu} f x_{1-\tau} = 0 \quad \text{Leb.} - a.s., \tag{2.5.26}$$

since $\sigma \frac{\partial^2}{\partial^2 \mu} f = \frac{\partial}{\partial \sigma} f$ holds for the normal distribution. From Lemma 2.6, to be shown, it follows that

$$\begin{aligned} b_2 + \frac{b_3}{\sigma} x_{1-j_0}^2 &= 0 \text{ Leb.} - a.s., \\ b_1 + \sum_{\tau=1}^{p} b_{\tau+3} x_{1-\tau} &= 0 \text{ Leb.} - a.s., \end{aligned}$$

so that $b_2 = b_3 = 0$ and $b_1 = b_4 = \ldots = b_{p+3} = 0$. □

Proof of Lemma 2.5. Let $\mu(\zeta, \phi_1, \ldots, \phi_p; x_0^p) = \zeta + \phi_1 x_0 + \ldots + \phi_p x_{1-p}$. Then

$$\begin{aligned} U_1^{\zeta} &= \frac{\frac{\partial}{\partial \mu} f(X_1; \mu(\zeta, \phi_1, \ldots, \phi_p; X_0^p), \sigma)}{f(X_1; \mu(\zeta, \phi_1, \ldots, \phi_p; X_0^p), \sigma)}, \\ U_1^{\phi_\tau} &= \frac{\frac{\partial}{\partial \mu} f(X_1; \mu(\zeta, \phi_1, \ldots, \phi_p; X_0^p), \sigma) X_{1-\tau}}{f(X_1; \mu(\zeta, \phi_1, \ldots, \phi_p; X_0^p), \sigma)} = U_1^{\zeta} X_{1-\tau}, \quad \tau = 1, \ldots, p, \\ Y_1 &= \frac{\frac{\partial}{\partial \sigma} f(X_1; \mu(\zeta, \phi_1, \ldots, \phi_p; X_0^p), \sigma)}{f(X_1; \mu(\zeta, \phi_1, \ldots, \phi_p; X_0^p), \sigma)}, \\ Z_1 &= \frac{\frac{\partial^2}{\partial^2 \sigma} f(X_1; \mu(\zeta, \phi_1, \ldots, \phi_p; X_0^p), \sigma)}{f(X_1; \mu(\zeta, \phi_1, \ldots, \phi_p; X_0^p), \sigma)}. \end{aligned}$$

The covariance matrix of $(U_1^{\zeta}, U_1^{\phi_1}, \ldots, U_1^{\phi_p}, Y_1, Z_1)$ is non-degenerate if and only if these random variables are linearly independent (in L_2). Therefore, suppose that for some constants b_j

$$b_1 Z_1 + b_2 Y_1 + b_3 U_1^{\zeta} + \sum_{\tau=1}^{p} b_{3+\tau} U_1^{\phi_\tau} = 0 \quad a.s. \tag{2.5.27}$$

holds. Since the distribution of $X_1, \ldots, X_{1-p}$ is equivalent to Lebesgue measure on $\mathbb{R}^{p+1}$, (2.5.27) is equivalent to

$$b_1 \frac{\partial}{\partial^2 \sigma} f + b_2 \frac{\partial}{\partial \sigma} f + b_3 \frac{\partial}{\partial \mu} f + \sum_{\tau=1}^{p} b_{3+\tau} \frac{\partial}{\partial \mu} f x_{1-\tau} = 0 \quad \text{Leb.} - a.s. \tag{2.5.28}$$

with $f = f(x_1; \mu(\zeta, \phi_1, \dots, \phi_p; x_0^p), \sigma)$. From (2.3.6) it follows that $b_1 = b_2 = 0$ and

$$b_3 + \sum_{\tau=1}^{p} b_{3+\tau} x_{1-\tau} = 0 \quad \text{Leb.} - a.s.,$$

so that $b_3 = \ldots = b_{3+p} = 0$. □

Proof of Lemma 2.6. The characteristic function of a normally distributed random variable with expectation μ and standard deviation $\sigma > 0$ is

$$\varphi(t; \mu, \sigma) = \exp\big(it\mu - \frac{\sigma^2 t^2}{2}\big).$$

The partial derivatives are

$$\begin{aligned} \frac{\partial \varphi(t; \mu, \sigma)}{\partial \sigma} &= -\sigma t^2 \varphi(t; \mu, \sigma), \\ \frac{\partial^2 \varphi(t; \mu, \sigma)}{\partial^2 \sigma} &= (\sigma^2 t^4 - t^2) \varphi(t; \mu, \sigma) \end{aligned}$$

and

$$\frac{\partial \varphi(t; \mu, \sigma)}{\partial \mu} = it\varphi(t; \mu, \sigma).$$

Taking the Fourier transform in (2.3.7) and interchanging integral and derivative gives

$$a_1 \frac{\partial \varphi(t; \mu, \sigma)}{\partial \mu} + a_2 \frac{\partial \varphi(t; \mu, \sigma)}{\partial \sigma} + a_3 \frac{\partial^2 \varphi(t; \mu, \sigma)}{\partial^2 \sigma} = 0 \quad \text{for all } t \in \mathbb{R}. \tag{2.5.29}$$

Plugging the partial derivatives into (2.5.29) and dividing by $\varphi(t; \mu, \sigma)$ yields

$$a_1 it - a_2 \sigma t^2 + a_3(\sigma^2 t^4 - t^2) = 0 \quad \text{for all } t \in \mathbb{R}$$

which is equivalent to

$$a_1 it + (-a_2 \sigma - a_3) t^2 + a_3 \sigma^2 t^4 = 0 \quad \text{for all } t \in \mathbb{R}. \tag{2.5.30}$$

Plugging in $t = -1$ and $t = 1$ in (2.5.30) and subtracting, we get $a_1 = 0$. Since the monomials $\{1, t, \dots, t^4\}$ build a basis of the vector space of all polynomials of degree less than or equal to 4, we get $a_2 = a_3 = 0$. The result follows by the inversion formula for probability density functions (see e.g. Billingsley, 1995). □

Proof of Lemma 2.8. Let $\sigma(\vartheta, \phi_1, \dots, \phi_p; x_0^p) = \big(\vartheta + \phi_1 x_0^2 + \ldots + \phi_d x_{1-p}^2\big)^{(1/2)}$. Then setting

$\sigma = \sigma(\vartheta, \phi_1, \ldots, \phi_p; X_0^p)$,

$$\begin{aligned} U_1^{\phi_j} &= \frac{\frac{\partial}{\partial\sigma} f\big(X_1;\sigma\big) X_{1-j}^2}{f\big(X_1;\sigma\big)2\sigma}, \quad j = 1, \ldots, p, \\ Y_1 &= \frac{\frac{\partial}{\partial\sigma} f\big(X_1;\sigma\big)/(2\sigma)}{f\big(X_1;\sigma\big)}, \\ Z_1 &= \frac{\frac{\partial^2}{\partial^2\sigma} f\big(X_1;\sigma\big)/(4\sigma^2) - \frac{\partial}{\partial\sigma} f\big(X_1;\sigma\big)/(4\sigma^3)}{f\big(X_1;\sigma\big)}. \end{aligned}$$

Again, the covariance is non-degenerate if and only if these random variables are linearly independent in L_2. Therefore, suppose that for constants b_j,

$$b_1 Z_1 + b_2 Y_1 + \sum_{j=1}^{p} b_{j+2} U_1^{\phi_j} = 0 \quad a.s. \tag{2.5.31}$$

Again, (2.5.31) is equivalent to

$$b_1\Big(\frac{\partial^2}{\partial^2\sigma} f/(2\sigma) - \frac{\partial}{\partial\sigma} f/(2\sigma)\Big) + b_2 \frac{\partial}{\partial\sigma} f + \sum_{j=1}^{p} b_{j+2} \frac{\partial}{\partial\sigma} f x_{1-j}^2 = 0 \quad \text{Leb.} - a.s.$$

where $f = f\big(x_1; \sigma(\vartheta, \phi_1, \ldots, \phi_p; x_0^p)\big)$. From (2.3.9), $b_1 = 0$ (as coefficient of $\frac{\partial^2}{\partial^2\sigma} f$) and

$$b_2 + \sum_{j=1}^{p} b_{j+2} x_{1-j}^2 = 0 \quad \text{Leb.} - a.s.,$$

so that $b_2 = b_3 = \ldots = b_{p+2} = 0$. □

Proof of Theorem 2.3. It is clear that

$$EM_n^{(K)} \le M_n \le \frac{\big((\sum \tilde{Z}_i)^+\big)^2}{\sum \tilde{Z}_i^2} + o_P(1).$$

Since one of the starting values in the EM-test is assumed to be $\alpha_J = 1/2$ and since the EM algorithm only increases the value of the likelihood (even though applied to a penalized quasi likelihood, see below for the argument), using the same argument as in the end of the proof of Theorem 2.2, we have

$$EM_n^{(K)} \ge \frac{\big((\sum \tilde{Z}_i)^+\big)^2}{\sum \tilde{Z}_i^2} + o_P(1)$$

and the claim follows. □

Derivation of the EM Property. For the argument, given the sample $X_1 = x_1, \dots, X_n = x_n$, we work with a (hypothetic) independent regime $(S_k)_{k\geq 0}$. The parameter vector is then given by $\boldsymbol{\psi}^T = (\alpha, \vartheta_1, \vartheta_2, \boldsymbol{\eta}^T) \in \mathbb{R}^{d+3}$, where α is the probability for state 2 for the independent regime. Let

(i) $\mathbf{S} = (S_1, \dots, S_n)$, $\mathbf{X} = (X_1, \dots, X_n)$, $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{s} = (s_1, \dots, s_n)$,

(ii) q be the joint pdf of $(\mathbf{X}, \mathbf{S})$ given $X_0^p, \boldsymbol{\psi}$ (under this artificial model),

(iii) r be the pdf of $\mathbf{S}$ given $\mathbf{X}, X_0^p, \boldsymbol{\psi}$ (also under this artificial model).

so that

$$p(\mathbf{x}|x_0^p, \boldsymbol{\psi}) r(\mathbf{s}|\mathbf{x}, x_0^p, \boldsymbol{\psi}) = q(\mathbf{x}, \mathbf{s}|x_0^p, \boldsymbol{\psi}). \tag{2.5.32}$$

Explicitly, we have

$$\begin{aligned}
p(\mathbf{x}|x_0^p, \boldsymbol{\psi}) &= \prod_{k=1}^{n} \{(1-\alpha) g(x_k|x_{k-1}^p; \vartheta_1, \boldsymbol{\eta}) + \alpha g(x_k|x_{k-1}^p; \vartheta_2, \boldsymbol{\eta})\}, \\
r(\mathbf{s}|\mathbf{x}, x_0^p, \boldsymbol{\psi}) &= \prod_{k=1}^{n} \frac{(1-\alpha)^{\mathbb{1}\{s_k=1\}} \alpha^{\mathbb{1}\{s_k=2\}} g(x_k|x_{k-1}^p; \vartheta_{s_k}, \boldsymbol{\eta})}{(1-\alpha) g(x_k|x_{k-1}^p; \vartheta_1, \boldsymbol{\eta}) + \alpha g(x_k|x_{k-1}^p; \vartheta_2, \boldsymbol{\eta})}, \\
q(\mathbf{x}, \mathbf{s}|x_0^p, \boldsymbol{\psi}) &= \prod_{k=1}^{n} (1-\alpha)^{\mathbb{1}\{s_k=1\}} \alpha^{\mathbb{1}\{s_k=2\}} g(x_k|x_{k-1}^p; \vartheta_{s_k}, \boldsymbol{\eta}).
\end{aligned}$$

Denote by $E_{\boldsymbol{\psi}^{(k)}}$ expectation w.r.t. the (artificial) distribution including the independent regime under the parameter $\boldsymbol{\psi}^{(k)}$. From (2.5.32), we get

$$pl_n(\boldsymbol{\psi}) = Q(\boldsymbol{\psi}|\boldsymbol{\psi}^{(k)}) - R(\boldsymbol{\psi}|\boldsymbol{\psi}^{(k)}) + p(\alpha),$$

where

$$\begin{aligned}
\bar{Q}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(k)}) &= E_{\boldsymbol{\psi}^{(k)}} \big(\log(q(\mathbf{X}, \mathbf{S}|X_0^p, \boldsymbol{\psi})) | \mathbf{X}, X_0^p, \boldsymbol{\psi}^{(k)} \big) + p(\alpha), \\
R(\boldsymbol{\psi}|\boldsymbol{\psi}^{(k)}) &= E_{\boldsymbol{\psi}^{(k)}} \{ \log(r(\mathbf{S}|\mathbf{X}, X_0^p, \boldsymbol{\psi})) | \mathbf{X}, X_0^p, \boldsymbol{\psi}^{(k)} \}
\end{aligned}$$

and $\boldsymbol{\psi}^{(k)}$ is the current value of $\boldsymbol{\psi}$. Then

$$\bar{Q}(\boldsymbol{\psi}^{(k+1)}|\boldsymbol{\psi}^{(k)}) \geq \bar{Q}(\boldsymbol{\psi}^{(k)}|\boldsymbol{\psi}^{(k)}) \Longrightarrow pl_n(\boldsymbol{\psi}^{(k+1)}) \geq pl_n(\boldsymbol{\psi}^{(k)}). \tag{2.5.33}$$

Proof of (2.5.33). Using Jensen's inequality we get:

$$R(\boldsymbol{\psi}^{(k+1)}|\boldsymbol{\psi}^{(k)}) - R(\boldsymbol{\psi}^{(k)}|\boldsymbol{\psi}^{(k)}) = E_{\boldsymbol{\psi}^{(k)}} \left\{ \log \frac{r(\mathbf{S}|\mathbf{X}, X_0^p, \boldsymbol{\psi}^{(k+1)})}{r(\mathbf{S}|\mathbf{X}, X_0^p, \boldsymbol{\psi}^{(k)})} \middle| \mathbf{X}, X_0^p, \boldsymbol{\psi}^{(k)} \right\}$$

$$\begin{aligned}
&\leq \log E_{\boldsymbol{\psi}^{(k)}} \left\{ \frac{r(\mathbf{S}|\mathbf{X}, X_0^p, \boldsymbol{\psi}^{(k+1)})}{r(\mathbf{S}|\mathbf{X}, X_0^p, \boldsymbol{\psi}^{(k)})} \middle| \mathbf{X}, X_0^p, \boldsymbol{\psi}^{(k)} \right\} \\
&= 0,
\end{aligned}$$

and therefore

$$\begin{aligned}
pl_n(\boldsymbol{\psi}^{(k)}) &= \bar{Q}(\boldsymbol{\psi}^{(k)}|\boldsymbol{\psi}^{(k)}) - R(\boldsymbol{\psi}^{(k)}|\boldsymbol{\psi}^{(k)}) \\
&\leq \bar{Q}(\boldsymbol{\psi}^{(k+1)}|\boldsymbol{\psi}^{(k)}) - R(\boldsymbol{\psi}^{(k)}|\boldsymbol{\psi}^{(k)}) \\
&\leq \bar{Q}(\boldsymbol{\psi}^{(k+1)}|\boldsymbol{\psi}^{(k)}) - R(\boldsymbol{\psi}^{(k+1)}|\boldsymbol{\psi}^{(k)}) \\
&= pl_n(\boldsymbol{\psi}^{(k+1)}).
\end{aligned}$$

□

Next we show that $\bar{Q}(\boldsymbol{\psi}^{(k+1)}|\boldsymbol{\psi}^{(k)}) \geq \bar{Q}(\boldsymbol{\psi}^{(k)}|\boldsymbol{\psi}^{(k)})$ holds for the updates obtained by the ECM algorithm (as proposed in Meng and Rubin, 1993). Relabel $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_{d+3})$ and $1 \leq r \leq d+3$ let

$$\pi_{\{t_1,\ldots,t_r\}} : \mathbb{R}^{d+3} \to \mathbb{R}^r, \quad \pi_{\{t_1,\ldots,t_r\}}(\psi_1, \ldots, \psi_{d+3}) = (\psi_{t_1}, \ldots, \psi_{t_r}),$$

$P_1, \ldots, P_q$ any partition of $\{1, \ldots, d+3\}$ and $-P_j = \{1, \ldots, d+3\} \setminus P_j$.
The ECM algorithm proceeds as follows.
Step 1: Compute $\boldsymbol{\psi}^{(k+1/q)} = \arg\max_{\boldsymbol{\psi}} \bar{Q}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(k)})$ subject to $\pi_{-P_1}(\boldsymbol{\psi}) = \pi_{-P_1}(\boldsymbol{\psi}^{(k)})$.
Step 2: Compute $\boldsymbol{\psi}^{(k+2/q)} = \arg\max_{\boldsymbol{\psi}} \bar{Q}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(k)})$ subject to $\pi_{-P_2}(\boldsymbol{\psi}) = \pi_{-P_2}(\boldsymbol{\psi}^{(k+1/q)})$.
⋮
Step q: Compute $\boldsymbol{\psi}^{(k+q/q)} = \arg\max_{\boldsymbol{\psi}} \bar{Q}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(k)})$ subject to $\pi_{-P_q}(\boldsymbol{\psi}) = \pi_{-P_q}(\boldsymbol{\psi}^{(k+(q-1)/q)})$.
The updated value is given by $\boldsymbol{\psi}^{(k+1)} = \boldsymbol{\psi}^{(k+q/q)}$. Then, by construction, we have

$$\bar{Q}(\boldsymbol{\psi}^{(k+1)}|\boldsymbol{\psi}^{(k)}) \geq \bar{Q}(\boldsymbol{\psi}^{(k+(q-1)/(q))}|\boldsymbol{\psi}^{(k)}) \geq \ldots \geq \bar{Q}(\boldsymbol{\psi}^{(k+1/q)}|\boldsymbol{\psi}^{(k)}) \geq \bar{Q}(\boldsymbol{\psi}^{(k)}|\boldsymbol{\psi}^{(k)}).$$

which implies (2.5.33).

Since

$$\begin{aligned}
&\bar{Q}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(k)}) \\
&= \sum_{i=1}^{n} \{\log((1-\alpha) g(X_i|X_{i-1}^p; \vartheta_1, \boldsymbol{\eta}))(1 - w_i^{(k)}) + \log(\alpha g(X_i|X_{i-1}^p; \vartheta_2, \boldsymbol{\eta})) w_i^{(k)}\} + p(\alpha)
\end{aligned}$$

with

$$w_i^{(k)} = \frac{\alpha^{(k)} g(X_i|X_{i-1}^p; \vartheta_2^{(k)}, \boldsymbol{\eta}^{(k)})}{(1-\alpha^{(k)}) g(X_i|X_{i-1}^p; \vartheta_1^{(k)}, \boldsymbol{\eta}^{(k)}) + \alpha^{(k)} g(X_i|X_{i-1}^p; \vartheta_2^{(k)}, \boldsymbol{\eta}^{(k)})},$$

the algorithm in our EM-test is the ECM algorithm with $P_1 = \{\alpha, \vartheta_1, \vartheta_2\}$ and $P_2 = \{\boldsymbol{\eta}\}$.

3 Testing in a linear switching autoregressive model with normal innovations

In this chapter we discuss testing for homogeneity in a linear switching autoregressive model with possibly switching intercept under the alternative and normal innovations. Even for HMMs with state-dependent distributions $P(X_t \leq x|S_t = i) = \Phi\big((x - \zeta_i)/\sigma\big)$ no feasible methods for testing the hypothesis $H : m = 1$ against $K : m \geq 2$ have been available, yet (see e.g. Piger 2009). For mixture models Chen and Li (2009) recently developed the so called EM-test for testing for homogeneity in a normal mixture model in the presence of a structural parameter. They show that the asymptotic distribution of the EM-test statistic is a simple function of the $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ and χ_1^2 distributions.

3.1 Example 2.1.1 (reconsidered)

As noted in the previous chapter, the LRT for testing for homogeneity in a switching autoregressive model does not admit a usual χ^2 distribution, since parameters of the full model are not identifiable under the hypothesis. For example, the hypothesis of a single regime in model (3.1.1), i.e. $\mathcal{M} = \{1\}$, can be represented by $H : \zeta_1 = \zeta_2$ under which the parameters a_{12} and a_{21} are not identifiable. In addition to that, testing for homogeneity in the model

$$X_t = \zeta_{S_t} + \sum_{j=1}^{p} \phi_j X_{t-j} + \sigma\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0,1) \tag{3.1.1}$$

is much more involved since $\sigma\frac{\partial^2 f(x;\mu,\sigma)}{\partial^2\mu} = \frac{\partial f(x;\mu,\sigma)}{\partial\sigma}$ holds for the normal distribution. Therefore, Assumption 2.6 is not satisfied for model (3.1.1) and the previously introduced MQLRT for testing for homogeneity does not admit a simple $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ distribution. This problem also arises in the related problem of testing for homogeneity in homoscedastic two-component normal mixtures which is a special case of our problem letting $p = 0$ and $S_t \overset{iid}{\sim} Mult\big(1; (1-\alpha, \alpha)\big)$ and has been studied extensively by Chen and Chen (2003), Qin and Smith (2004) and Chen and Li (2009). Chen and Chen (2003) derive an asymptotic upper bound for the MLRT for testing for homogeneity in normal mixture models in the presence of a structural parameter which is strengthened by Qin and Smith (2004). They give a stochastic upper bound for the MLRT which has $\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$ distribution. But it is not clear at all whether this upper bound is also attained. Chen and Li (2009) investigate an EM-test for testing for homogeneity in this model.

In the following we extend the EM-test of Chen and Li (2009) to linear switching autoregressive models with possibly switching intercept under the alternative and normal innovations. To this end, we suppose that under the null hypothesis, i.e. no regime switch, $(X_k)_k$ is a causal AR(p) process. This assumption assures that the order p as well as the parameters of the autoregressive process are uniquely defined, cf. Kreiss and Neuhaus (2006). Throughout this chapter we assume $\sigma \in [\delta, \infty)$, $\delta > 0$, and $\zeta \in \Theta$ and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^T \in \boldsymbol{H}$, where Θ and $\boldsymbol{H}$ are any subsets of $\mathbb{R}$ and $\mathbb{R}^p$, respectively.

3.1.1 Penalized maximum likelihood

As in Chapter 2, following Cho and White (2007), we consider a model which ignores the serial correlation in $(S_k)_k$ but captures the serial correlation of the process $(X_k)_k$. Even if we ignore the serial correlation in $(S_k)_k$ we are able to test for the number of regimes. Let $X_1, \ldots, X_n$ be a random sample of size n from model (3.1.1). We do not work with the (full) likelihood conditional on the intial observations $(X_0, \ldots, X_{-p+1})$ and the initial state $S_0 = i_0$. Instead, we consider the *quasi log-likelihood function* which is given by

$$l_n(\boldsymbol{\psi}) = \sum_{t=1}^{n} \log\left((1-\alpha)g(X_t|X_{t-1}^p; \zeta_1, \boldsymbol{\phi}, \sigma) + \alpha g(X_t|X_{t-1}^p; \zeta_2, \boldsymbol{\phi}, \sigma)\right), \tag{3.1.2}$$

with $\boldsymbol{\psi} = (\alpha, \zeta_1, \zeta_2, \boldsymbol{\phi}^T, \sigma)^T$. Here, the parameter $(1-\alpha, \alpha)$ corresponds to the stationary distribution of the hidden Markov chain $(S_k)_k$, cf. Remark 1.1. Since we assume that the innovations $(\epsilon_k)_k$ are independent and identically normally distributed with expectation 0 and scale parameter σ, the conditional density (w.r.t. Lebesgue measure on $\mathbb{R}$) of X_t given $X_{t-1}^p = x_{t-1}^p$ and $S_t = i$ is

$$g(x_t|x_{t-1}^p; \zeta_i, \boldsymbol{\phi}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_t - \zeta_i - \sum_{j=1}^p \phi_j x_{t-j})^2}{2\sigma^2}\right).$$

In the next section, we give a test for testing the hypothesis of no regime switch, i.e.

$$H : \alpha(1-\alpha)(\zeta_1 - \zeta_2) = 0$$

in model (3.1.1) using a penalized version of (3.1.2).

3.2 The EM-test

Similar to Chen et al. (2001, 2004) and Chen and Li (2009) we consider the modified quasi log likelihood function defined by

$$pl_n(\alpha, \zeta_1, \zeta_2, \boldsymbol{\phi}, \sigma) = l_n(\alpha, \zeta_1, \zeta_2, \boldsymbol{\phi}, \sigma) + p(\alpha)$$

where $p(\alpha)$ is a penalty function fulfilling the properties given in Section 1.5.

As we know from Chen and Li (2009), the asymptotic distribution of the MLRT for testing for homogeneity in homoscedastic normal mixture models is still under investigation. In the following we give a (quasi) EM-test for testing the hypothesis of one regime against the alternative of (at least) two regimes in model (3.1.1).

We describe the EM-test in form of the following algorithm. Note that in this algorithm we apply some steps of the ECM algorithm instead of the EM algorithm. If we used some steps of the EM algorithm, we would have to derive the updated estimators $(\zeta_{1j}^{(k+1)}, \zeta_{2j}^{(k+1)}, \boldsymbol{\phi}_j^{(k+1)}, \sigma_j^{(k+1)})$ in Step 3 by maximizing

$$\sum_{t=1}^{n}(1 - w_{tj}^{(k)}) \log g(X_t|X_{t-1}^p; \zeta_1, \boldsymbol{\phi}, \sigma) + \sum_{t=1}^{n} w_{tj}^{(k)} \log g(X_t|X_{t-1}^p; \zeta_2, \boldsymbol{\phi}, \sigma)$$

simultaneously over $\Theta^2 \times \boldsymbol{H} \times [\delta, \infty)$. As shown in Section 2.5 (*Derivation of the EM property*) the test statistic increases with every iteration even if we perform the corresponding steps of the ECM algorithm.

Step 0. Choose $0 < \alpha_1 < \alpha_2 < \ldots < \alpha_J = 0.5$. Compute

$$(\widehat{\zeta}_0, \widehat{\boldsymbol{\phi}}_0, \widehat{\sigma}_0) = \arg\max_{\zeta, \boldsymbol{\phi}, \sigma} pl_n(0.5, \zeta, \zeta, \boldsymbol{\phi}, \sigma).$$

Put $j = 1$ and $k = 0$.

Step 1. Put $\alpha_j^{(k)} = \alpha_j$.

Step 2. Compute

$$(\zeta_{1j}^{(k)}, \zeta_{2j}^{(k)}, \boldsymbol{\phi}_j^{(k)}, \sigma_j^{(k)}) = \arg \max_{\zeta_1, \zeta_2, \boldsymbol{\phi}, \sigma} pl_n(\alpha_j^{(k)}, \zeta_1, \zeta_2, \boldsymbol{\phi}, \sigma)$$

and

$$M_n^{(k)}(\alpha_j) = 2\left\{pl_n(\alpha_j^{(k)}, \zeta_{1j}^{(k)}, \zeta_{2j}^{(k)}, \boldsymbol{\phi}_j^{(k)}, \sigma_j^{(k)}) - pl_n(0.5, \widehat{\zeta}_0, \widehat{\zeta}_0, \widehat{\boldsymbol{\phi}}_0, \widehat{\sigma}_0)\right\}.$$

Step 3. Compute for $t = 1, \ldots, n$ the weights

$$w_{tj}^{(k)} = \frac{\alpha_j^{(k)} g(X_t|X_{t-1}^p; \zeta_{2j}^{(k)}, \boldsymbol{\phi}_j^{(k)}, \sigma_j^{(k)})}{(1-\alpha_j^{(k)}) g(X_t|X_{t-1}^p; \zeta_{1j}^{(k)}, \boldsymbol{\phi}_j^{(k)}, \sigma_j^{(k)}) + \alpha_j^{(k)} g(X_t|X_{t-1}^p; \zeta_{2j}^{(k)}, \boldsymbol{\phi}_j^{(k)}, \sigma_j^{(k)})}.$$

Compute the estimators

$$\begin{aligned}
\alpha_j^{(k+1)} &= \arg\max_{\alpha} \left((n - \sum_{t=1}^n w_{tj}^{(k)}) \log(1-\alpha) + \sum_{t=1}^n w_{tj}^{(k)} \log(\alpha) + p(\alpha) \right) \\
\zeta_{1j}^{(k+1)} &= \frac{\sum_{t=1}^n (1 - w_{tj}^{(k)})(X_t - \sum_{\tau=1}^p \phi_{\tau j}^{(k)} X_{t-\tau})}{\sum_{t=1}^n (1 - w_{tj}^{(k)})} \\
\zeta_{2j}^{(k+1)} &= \frac{\sum_{t=1}^n w_{tj}^{(k)} (X_t - \sum_{\tau=1}^p \phi_{\tau j}^{(k)} X_{t-\tau})}{\sum_{t=1}^n w_{tj}^{(k)}} \\
\boldsymbol{\phi}_j^{(k+1)} &= \arg\max_{\boldsymbol{\phi}} \left(\sum_{t=1}^n (1 - w_{tj}^{(k)}) \log g(X_t|X_{t-1}^p; \zeta_{1j}^{(k+1)}, \boldsymbol{\phi}, \sigma_j^{(k)}) \right. \\
&\quad \left. + \sum_{t=1}^n w_{tj}^{(k)} \log g(X_t|X_{t-1}^p; \zeta_{2j}^{(k+1)}, \boldsymbol{\phi}, \sigma_j^{(k)}) \right) \\
\sigma_j^{(k+1)} &= \arg\max_{\sigma} \left(\sum_{t=1}^n (1 - w_{tj}^{(k)}) \log g(X_t|X_{t-1}^p; \zeta_{1j}^{(k+1)}, \boldsymbol{\phi}_j^{(k+1)}, \sigma) \right. \\
&\quad \left. + \sum_{t=1}^n w_{tj}^{(k)} \log g(X_t|X_{t-1}^p; \zeta_{2j}^{(k+1)}, \boldsymbol{\phi}_j^{(k+1)}, \sigma) \right).
\end{aligned}$$

Compute

$$M_n^{(k+1)}(\alpha_j) = 2 \left\{ pl_n(\alpha_j^{(k+1)}, \zeta_{1j}^{(k+1)}, \zeta_{2j}^{(k+1)}, \boldsymbol{\phi}_j^{(k+1)}, \sigma_j^{(k+1)}) - pl_n(0.5, \widehat{\zeta}_0, \widehat{\zeta}_0, \widehat{\boldsymbol{\phi}}_0, \widehat{\sigma}_0) \right\},$$

put $k = k+1$ and repeat Step 3 for a fixed number of iterations K.

Step 4. Put $j = j+1$, $k = 0$ and go to Step 1, until $j = J$.

Step 5. Compute the test statistic

$$EM_n^{(K)} = \max \left\{ M_n^{(K)}(\alpha_j), j = 1, \ldots, J \right\}.$$

Due to the construction of $EM_n^{(K)}$ we reject the null hypothesis of just a single regime, i.e. $\mathcal{M} = \{1\}$, when $EM_n^{(K)}$ exceeds some critical value which can be determined either via simulations or based on asymptotic results.

3.3 Asymptotics

Deriving the asymptotic distribution of the EM-test, as suggested in the previous section, is quite involved. Goffinet, Laurent and Loisel (1992) derive the asymptotic distribution of the LRT for homogeneity in homoscedastic normal mixture models if the proportion α is known a priori. They show that the limiting distribution of the LRT is a χ_1^2 distribution if $\alpha \neq 1/2$ and a $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ distribution if $\alpha = 1/2$. Based on this result, Chen and Li (2009) derive the asymptotic distribution of the EM-test for testing for homogeneity in normal mixture models in the presence of a structural parameter. The following theorem shows that the asymptotic distribution of the EM-test for testing the hypothesis of one regime against (at least) two regimes in model (3.1.1) is the same as for the EM-test, introduced by Chen and Li (2009) for testing for homogeneity in normal mixtures in the presence of the structural parameter σ.

Theorem 3.1. *Let $p(\alpha)$ be a continuous function such that $p(\alpha) \to -\infty$ as $\alpha \to 0$. Suppose that $p(\alpha)$ attains its maximal value at $1/2$. Provided that $\alpha_1 \neq 1/2$ and $1/2$ are contained in the set of initial values, we have under the null model and for every fixed K,*

$$P(EM_n^{(K)} \leq x) \longrightarrow F(x - \Delta)\big(\mathbb{1}_{\{x \geq 0\}} + F(x)\big)/2,\ n \to \infty,$$

where $F(\cdot)$ is the cdf of a χ_1^2-variate and

$$\Delta = 2 \max_{\alpha_j \neq 0.5} \{p(\alpha_j) - p(0.5)\}.$$

The proof of Thereom 3.1 is deferred to Section 3.6.

Remark 3.1. 1. Starting with just one initial value $\alpha_1 = 1/2$, the asymptotic distribution of the EM-test is a $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ distribution.
2. If $1/2 \notin \mathcal{J} = \{\alpha_1, \ldots, \alpha_J\}$, the asymptotic distribution of the EM-test is a shifted χ_1^2 distribution where the shift is due to the penalty function $p(\alpha)$.
3. Note that our results also cover the problem arising in the survey article of Piger (2009), who wants to test for one against two regimes in a normal hidden Markov model with state dependent distributions $P(X_t \leq x | S_t = i) = \Phi\big((x - \zeta_i)/\sigma\big)$, $i = 1, 2$ and $\Phi(\cdot)$ being the cdf of a standard normal variate.

If there is a priori knowledge of the proportion α, say $\alpha = 1/2$, a test based on this fixed proportion would be appropriate. For data not being from an alternative model with $\alpha = 1/2$, the choice of fixed $\alpha = 1/2$ under the alternative leads to a loss of power. Therefore, a test based on a set of fixed proportions, say $\mathcal{J} = \{\alpha_1, \ldots, \alpha_J\}$, would be desirable. As can be seen in the simulation section (see Section 3.4) the EM-steps do not (significantly) improve the power of our EM-test. So, we also suggest a test based on fixed proportions. Clearly a test based on $EM_n^{(0)}$ will be a test based on fixed proportions since $\alpha_j^{(0)} = \alpha_j$, $j = 1, \ldots, J$, and its asymptotics are known from Theorem 3.1. But using fixed proportions, we are not concerned with problems caused by non-identifiability problems (as long as $\{0,1\} \cap \mathcal{J} = \emptyset$). Therefore, a penalty function on α is not necessary. To this end, we define

$$R_n(\alpha_j) = 2\{l_n(\alpha_j, \widehat{\zeta}_{1,\alpha_j}, \widehat{\zeta}_{2,\alpha_j}, \widehat{\boldsymbol{\phi}}_{\alpha_j}, \widehat{\sigma}_{\alpha_j}) - l_n(0.5, \widehat{\zeta}_0, \widehat{\zeta}_0, \widehat{\boldsymbol{\phi}}_0, \widehat{\sigma}_0)\}, \quad \alpha_j \in \mathcal{J}.$$

Here, $(\widehat{\zeta}_{1,\alpha_j}, \widehat{\zeta}_{2,\alpha_j}, \widehat{\boldsymbol{\phi}}_{\alpha_j}, \widehat{\sigma}_{\alpha_j})$ is the maximizer of $l_n(\alpha, \zeta_1, \zeta_2, \boldsymbol{\phi}, \sigma)$ subject to $\alpha = \alpha_j$. Further, let

$$R_n(\mathcal{J}) = \max_{\alpha_j \in \mathcal{J}} R_n(\alpha_j).$$

The test for homogeneity based on fixed proportions in model (3.1.1) rejects H for large values of $R_n(\mathcal{J})$. In the following we give the asymptotic distribution of its test statistic.

Corollary 3.2. *Under the null model and whenever $\{\alpha_1, 0.5\} \subset \mathcal{J}$, with $\alpha_1 \neq 0.5$, we have*

$$P(R_n(\mathcal{J}) \leq x) \to F(x)\big(\mathbb{1}_{\{x \geq 0\}} + F(x)\big)/2,\ n \to \infty,$$

where $F(\cdot)$ is the cdf of a χ_1^2 variate.

Letting $p(\alpha) \equiv 0$ in the proof of Theorem 3.1 we see that the assumed limiting distribution in Corollary 3.2 serves as a stochastic upper bound for the test based on fixed proportions. Choosing the same values as in the end of the proof of Theorem 3.1 we see that this upper bound is also attained, asymptotically.

3.4 Simulations

In this section we present some of the results of an extensive simulation study of the EM-test and of the test based on fixed proportions. We choose $\mathcal{J} = \{0.1, 0.3, 0.5\}$ and $p(\alpha) = \log(1 - |1 - 2\alpha|)$ as penalty function for the EM-test. For the computation, we choose $\delta = 0.1$.

3.4.1 Simulated sizes

In the following we simulate the size of the EM-test and of the test based on fixed proportions for some data-generating processes (DGP):
DGP 1: $X_t = 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$.
Model 1: $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$.
The results for various sample sizes can be found in Table 3.1. Figure 3.2 shows the ecdf of the EM-test statistic $EM_n^{(2)}$ for sample size $n = 200$ as well as $n = 1000$.

Table 3.1: DGP: $X_t = 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, Model: $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$; number of replications: 20,000.

Sample Size	Nominal Levels (%)	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$R_n(\mathcal{J})$
$n = 200$	10%	12.3	12.8	13.1	15.7
	5%	6.7	7.1	7.3	8.4
	1%	1.5	1.7	1.8	1.9
$n = 500$	10%	11.9	12.0	12.2	14.9
	5%	6.1	6.2	6.4	8.0
	1%	1.2	1.3	1.3	1.8
$n = 1000$	10%	10.8	10.9	11.0	13.8
	5%	5.6	5.7	5.7	7.4
	1%	1.2	1.2	1.3	1.6

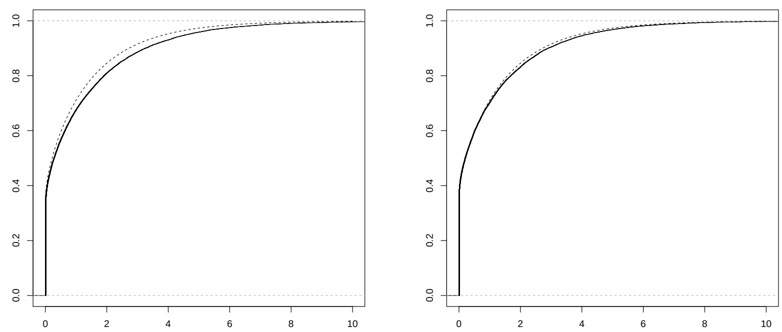

Figure 3.1: Ecdf of $EM_n^{(2)}$ for testing for homogeneity in model (3.1.1) (solid line) for DGP $X_t = 0.5X_{t-1}+\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, together with the limit distribution (dashed line) for $n = 200$ (left) and $n = 1000$ (right).

DGP 2: $X_t = -0.2X_{t-1} + 0.4X_{t-2} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$.
Model 2: $X_t = \zeta_{S_t} + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \sigma\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$.
The results for various sample sizes are in Table 3.2.

The EM-test is somewhat anticonservative for small sample sizes in both scenarios. Introducing a penalty function on σ which forces σ to be bounded away from zero will lead to more accurate type I errors of the EM-test. But introducing a penalty function on σ would be a tradeoff between more accurate type I errors and the EM-test having a significantly too large null proportion.

We also used other penalty functions on α, such as $\log(4\alpha(1-\alpha))$ or $3\log(1-|1-2\alpha|)$, since the penalty function influences the asymptotic distribution of the EM-test statistic via $\Delta = 2\max_{\alpha_j \neq 0.5}\{p(\alpha_j) - p(0.5)\}$. The results were essentially the same, though, so that we do not display the results, here.

Table 3.2: DGP: $X_t = -0.2X_{t-1} + 0.4X_{t-2} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, Model: $X_t = \zeta_{S_t} + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \sigma\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$; number of replications: 20,000.

Sample Size	Nominal Levels (%)	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$R_n(\mathcal{J})$
$n = 200$	10%	13.1	13.6	13.9	16.6
	5%	7.0	7.5	7.7	9.2
	1%	1.7	1.9	2.0	2.2
$n = 500$	10%	11.8	11.9	12.0	15.0
	5%	6.1	6.3	6.4	8.0
	1%	1.3	1.3	1.4	1.9
$n = 1000$	10%	10.9	11.0	11.1	13.9
	5%	5.6	5.6	5.7	7.2
	1%	1.3	1.3	1.3	1.6

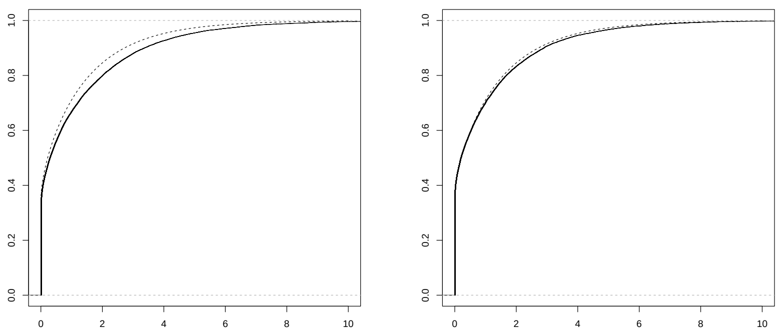

Figure 3.2: Ecdf of $EM_n^{(2)}$ for testing for homogeneity in model (3.1.1) (solid line) for DGP $X_t = -0.2X_{t-1} + 0.4X_{t-2} + \epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, together with the limit distribution (dashed line) for $n = 200$ (left) and $n = 1000$ (right).

3.4.2 Power comparison

We present the results of a power comparison between the EM-test, the test based on fixed proportions and the quasi likelihood ratio test (QLRT) by Cho and White (2007). To ensure fairness, we use simulated critical values in all scenarios for the three tests.
We choose the following data-generating processes: $X_t = (-1)^{S_t}\zeta + \phi X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, with various values of ζ, ϕ, a_{12} and a_{21}. For $\phi = 0.5$ and $a_{12} = a_{21}$, implying $\alpha = 0.5$, we can compare the results shown in Table 3.3 with the results given in Cho and White (2007, Section 3, Table 3). Using size distortion-adjusted critical values the results in Cho and White (2007) slightly differ from ours for the QLRT. The EM-test, the test based on fixed proportions and the QLRT have almost identical powers. In some scenarios the EM-test outperforms the QLRT and the test based on fixed proportions.

Table 3.3: Nominal level: 5%; DGP: $X_t = (-1)^{S_t}\zeta + 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, sample size: 500, number of replications: 5,000, Model: $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$, with $\epsilon_t \overset{iid}{\sim} N(0,1)$. Let $\alpha = a_{12}/(a_{12}+a_{21})$ and $(1-\alpha,\alpha)$ be the stationary distribution of the hidden Markov Chain $(S_k)_k$.

a_{12}	a_{21}	α	ζ	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	QLRT	$R_n(\mathcal{J})$
0.1	0.1	0.5	1.0	6.4	6.3	6.3	6.1	6.2
0.2	0.2	0.5	1.0	30.6	30.9	31.0	23.0	24.7
0.3	0.3	0.5	1.0	67.9	68.0	68.0	57.2	58.7
0.5	0.5	0.5	1.0	87.7	87.6	87.7	81.4	83.2
0.7	0.7	0.5	1.0	75.0	75.3	75.0	64.1	67.0
0.8	0.8	0.5	1.0	58.2	58.4	58.5	45.5	48.5
0.9	0.9	0.5	1.0	35.9	35.9	36.2	25.1	27.1

As can be seen in Tables 3.3 and 3.4 the power of the three tests does not only depend on the stationary distribution $(1-\alpha,\alpha)$ but also on the transition probabilities of the hidden Markov chain $(S_k)_k$. The tests have the highest power when the Markov chain reduces to an i.i.d. sample, i.e. $a_{12} = a_{21} = 0.5$.

Table 3.4: Nominal level: 5%; DGP: $X_t = (-1)^{S_t}\zeta + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, sample size: 500, number of replications: 5,000, Model: $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$, with $\epsilon_t \overset{iid}{\sim} N(0,1)$. Let $\alpha = a_{12}/(a_{12}+a_{21})$ and $(1-\alpha,\alpha)$ be the stationary distribution of the hidden Markov Chain $(S_k)_k$.

a_{12}	a_{21}	α	ζ	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	QLRT	$R_n(\mathcal{J})$
0.1	0.1	0.5	1.0	12.9	12.9	12.9	10.0	10.7
0.2	0.2	0.5	1.0	34.3	34.4	34.4	26.0	28.1
0.3	0.3	0.5	1.0	65.4	65.5	65.3	54.4	56.8
0.5	0.5	0.5	1.0	88.1	88.3	88.4	81.9	82.8
0.7	0.7	0.5	1.0	63.9	63.8	64.0	52.7	54.7
0.8	0.8	0.5	1.0	30.2	30.2	30.4	21.3	24.2
0.9	0.9	0.5	1.0	9.3	9.3	9.4	6.4	7.2

This is in sharp contrast to testing for homogeneity in a hidden Markov-model with state dependent distributions $P(X_t \leq x|S_t = i) = \Phi\big((x-\zeta_i)/\sigma\big)$, $i = 1,2$, using the EM-test introduced by Chen and Li (2009) and neglecting the dependence structure under the alternative, see Table 3.5. Testing for homogeneity in a Poisson-HMM via the MLRT based on the log-likelihood under independence assumption, Dannemann (2009) also shows via simulations that different transitions matrices leading to the same stationary distribution do not have much influence to the power of the used test.

Table 3.5: Nominal level: 5%; DGP: $X_t = (-1)^{S_t}\zeta + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, sample size: 500, number of replications: 5,000, Model: $X_t = \zeta_{S_t} + \sigma\epsilon_t$, with $\epsilon_t \overset{iid}{\sim} N(0,1)$. Let $\alpha = a_{12}/(a_{12}+a_{21})$ and $(1-\alpha,\alpha)$ be the stationary distribution of the hidden Markov Chain $(S_k)_k$.

a_{12}	a_{21}	α	ζ	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$
0.1	0.1	0.5	1.0	87.6	87.6	87.6
0.2	0.2	0.5	1.0	88.0	88.0	88.0
0.3	0.3	0.5	1.0	89.3	89.3	89.3
0.5	0.5	0.5	1.0	89.5	89.5	89.5
0.7	0.7	0.5	1.0	88.6	88.6	88.6
0.8	0.8	0.5	1.0	89.5	89.5	89.5
0.9	0.9	0.5	1.0	89.0	89.0	89.0

In Table 3.6 we report the results for DGP's $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$ for various combinations ζ_1, ζ_2, σ, a_{12} and a_{21} (leading to values $\alpha \neq 0.5$). It can be seen that the QLRT and the test based on fixed proportions perform slightly better than the EM-test for small values of α.

Table 3.6: Nominal level: 5%; DGP: $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, sample size: 500, number of replications: 5,000, Model: $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$, with $\epsilon_t \overset{iid}{\sim} N(0,1)$. Let $\alpha = a_{12}/(a_{12}+a_{21})$ and $(1-\alpha,\alpha)$ be the stationary distribution of the hidden Markov Chain $(S_k)_k$.

a_{12}	a_{21}	α	ζ_1	ζ_2	ϕ	σ	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	QLRT	$R_n(\mathcal{J})$
0.1	0.9	0.1	−1	1	0.5	1	77.4	77.5	77.5	84.1	86.1
0.05	0.45		−1	1	0.5	1	57.7	57.6	57.6	67.8	69.6
0.2	0.8	0.2	−1	1	0.5	1	93.5	93.5	93.5	94.3	95.1
0.15	0.6		−1	1	0.5	1	89.7	89.7	89.7	90.0	91.6
0.1	0.4		−1	1	0.5	1	72.6	72.7	72.6	75.2	78.0
0.3	0.7	0.3	−1	1	0.5	1	91.6	91.7	91.7	90.4	91.2
0.2	7/15		−1	1	0.5	1	85.6	85.9	86.0	82.8	84.7
0.15	0.35		−1	1	0.5	1	64.4	64.4	64.4	61.7	63.8
0.4	0.6	0.4	−1	1	0.5	1	89.2	89.6	89.7	85.7	85.8
0.2	0.3		−1	1	0.5	1	50.9	51.1	51.5	44.7	45.8

3.5 Application

In this section, we apply our methods to the series of quarterly, saisonally adjusted U.S. GNP from 1947(1) to 2002(3). The data are Real U.S. Gross National Product in billions of chained 1996 dollars and can be obtained by the Federal Reserve Bank of St. Louis (*http://research.stlouisfed.org/*). Instead of considering the data, say Y_t, we consider the growth rate $X_t = \nabla \log(Y_t) = \log(Y_t) - \log(Y_{t-1})$ (in %) which is plotted in Figure 3.3 (left).

Marginal distribution

To start, we investigate the marginal distribution of the data. Figure 3.3 (right) contains a histogram together with the density of a fitted two-component normal mixture and a kernel density estimate. We test one against two components in a normal mixture model

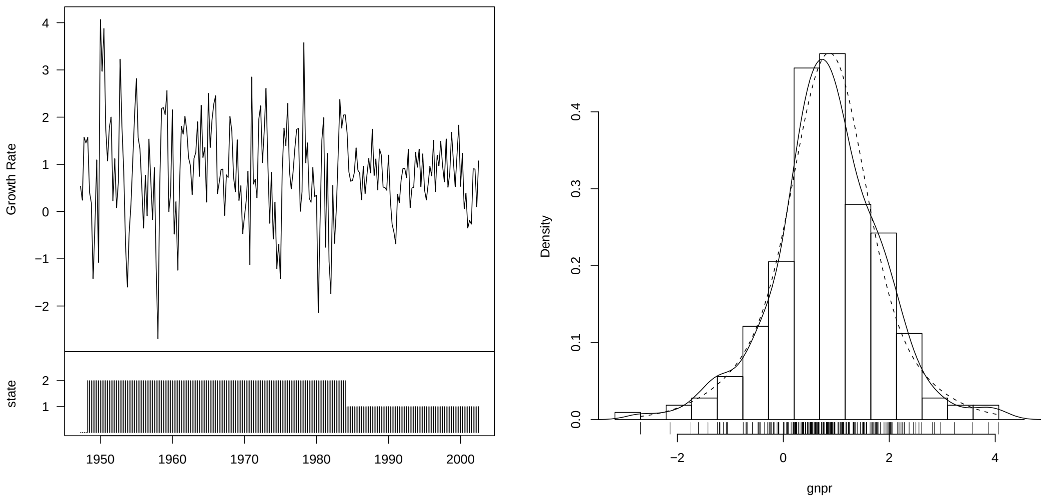

Figure 3.3: U.S. GNP quarterly growth rate in % (left) and histogram together with the density of a fitted two-component normal mixture (dashed line) and a kernel density estimate (dotted line) of the GNP growth rate (right). The vertical lines in the left plot indicate the states of the observations computed with the Viterbi algorithm given the fitted model $\mathcal{M}_3$ (with $p = 1$).

by using the test in Chen and Li (2009). If we test against an alternative mixture model with distinct means but equal variances, the hypothesis of a single component cannot be rejected (p-value $= 0.99$). If we test against an alternative with possibly distinct means and variances, the hypothesis is strongly rejected in favor of two components (p-value $= 0.0045$). It turns out that the alternative two-component model has almost equal means ($\overline{\mu}_1 = 0.79$ and $\overline{\mu}_2 = 0.87$) but quite distinct standard deviations ($\overline{\sigma}_1 = 1.29$, $\overline{\sigma}_2 = 0.62$). Further, if we test against the specific alternative model with equal means but distinct variances, the hypothesis of a single component is rejected as well (p-value $= 0.00043$).

Autoregressive model

As indicated by the acf and pacf (see Figure 3.4), the series shows autocorrelation. Therefore, we model the data by AR(p) models ($\mathcal{M}_1$), $p = 1, \ldots, 4$,

$$X_t = \zeta + \sum_{j=1}^{p} \phi_j X_{t-j} + \sigma \epsilon_t$$

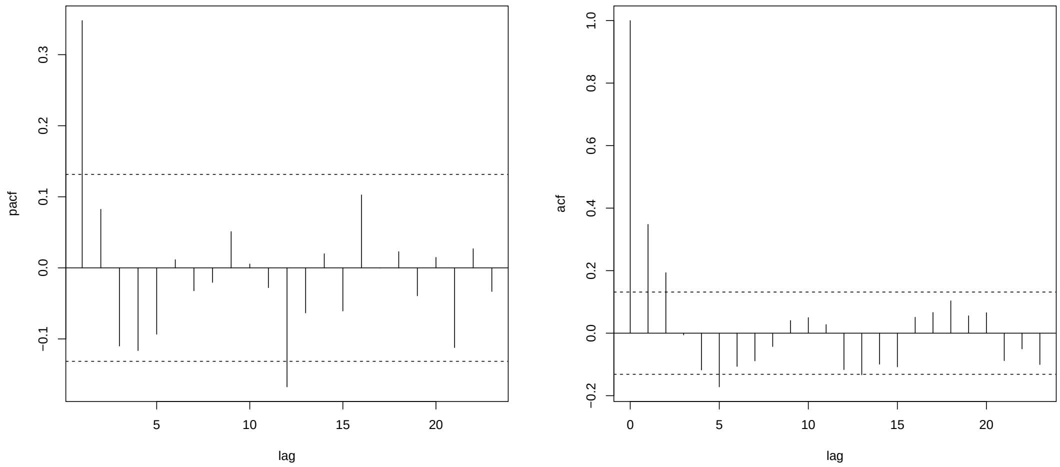

Figure 3.4: Sample pacf (left) and acf (right) of the U.S. GNP quarterly growth rate. The dashed line gives an approximate 95% confidence interval.

with normal innovations and several switching autoregressive models ($\mathcal{M}_2 \ldots$ totally switching, $\mathcal{M}_3 \ldots$ switching scale parameter (model (2.2.3)) and $\mathcal{M}_4 \ldots$ switching intercept, see model (3.1.1)). Using formal model selection criteria, one chooses model $\mathcal{M}_3$ and $p = 1$ according to BIC and $\mathcal{M}_3$ with $p = 3$ according to AIC. Here, we note that the AIC and BIC are computed by

$$\text{AIC} = -2\tilde{l}_n(\hat{\boldsymbol{\omega}}) + 2 \cdot k(\hat{\boldsymbol{\omega}}) \text{ and } \text{BIC} = -2\tilde{l}_n(\hat{\boldsymbol{\omega}}) + \log(n) \cdot k(\hat{\boldsymbol{\omega}}),$$

where $\tilde{l}_n(\cdot)$ is the full model log likelihood conditional on the first 4 observations and on state 1 and $k(\hat{\boldsymbol{\omega}})$ denotes the length of $\hat{\boldsymbol{\omega}}$. Results are shown in Table 3.7.

Table 3.7: BIC (left) and AIC (right) for the corresponding models for series 1947(1)–2002(3)

BIC	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	**AIC**	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
$p = 1$	615.05	602.91	**592.56**	622.51	$p = 1$	604.90	575.83	572.25	602.21
$p = 2$	618.86	607.16	593.71	626.57	$p = 2$	605.32	573.32	570.02	602.88
$p = 3$	621.54	615.63	597.01	628.85	$p = 3$	604.62	575.01	**569.94**	601.77
$p = 4$	623.91	623.73	600.71	632.39	$p = 4$	603.60	576.35	570.25	601.93

We cannot reject the hypothesis of a purely linear autoregressive model in favor of model $\mathcal{M}_4$ with two states, using the EM-test. Testing for homogeneity in model $\mathcal{M}_3$ via the MQLRT (introduced in Chapter 2), we are able to reject the hypothesis of no regime switch for $p = 1, \ldots, 4$, see Table 3.8. A similar model has been introduced by Bhar and Hamori (2004) who find evidence of two volatility states. But considering the ML estimates in model $\mathcal{M}_3$ (see Table 3.9), we see that the hidden Markov chain is highly persistent. Computing the most likely sequence of hidden states using the Viterbi algorithm

Table 3.8: Fits of mixture AR(p) models with possibly switching variance using modified quasi MLE with normal innovations and results of the MQLRT, series 1947(1)–2002(3).

	$\widehat{\alpha}^*$	$\widehat{\zeta}^*$	$\widehat{\phi}_1^*$	$\widehat{\phi}_2^*$	$\widehat{\phi}_3^*$	$\widehat{\phi}_4^*$	$\widehat{\sigma}_1^*$	$\widehat{\sigma}_2^*$	M_n	p-value
$p=1$	0.42	0.58	0.33				0.61	1.29	12.63	< 0.01
$p=2$	0.44	0.55	0.32	0.05			0.61	1.27	11.46	< 0.01
$p=3$	0.42	0.60	0.33	0.09	-0.13		0.61	1.28	12.88	< 0.01
$p=4$	0.45	0.69	0.29	0.12	-0.09	-0.15	0.56	1.26	15.17	< 0.01

Table 3.9: Fits for model $\mathcal{M}_3$ (only the variance is allowed to switch, using maximum likelihood estimation) 1947(1)–2002(3).

	$\hat{a}_{12}$	$\hat{a}_{21}$	$\hat{\zeta}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$\hat{\sigma}_1$	$\hat{\sigma}_2$
$p=1$	0.007	0.005	0.51	0.35				0.51	1.12
$p=2$	0.007	0.005	0.44	0.30	0.14			0.50	1.12
$p=3$	0.007	0.005	0.48	0.31	0.17	-0.10		0.50	1.12
$p=4$	0.007	0.005	0.53	0.31	0.19	-0.07	-0.09	0.50	1.11

(see Viterbi, 1967) given the fitted model $\mathcal{M}_3$ ($p = 1$), we see that there is only one regime switch in the variance between 1984(1) and 1984(2), see Figure 3.3 (left). This is a result of the great 'Great Moderation' of the U.S. GNP growth rate, i.e. the permanent decline in the growth rate of U.S. GNP. Therefore, we divide our series in two subseries: the first from 1947(1)–1984(1) and the second 1984(2)–2002(3).
Again, we fit several models $\mathcal{M}_1, \dots, \mathcal{M}_4$ to the two subseries. For the series from 1947(1)–1984(1), the BIC, see Table 3.10, as well as the AIC favor a purely autoregressive model of order 1, see Table 3.11. Testing for homogeneity in model $\mathcal{M}_4$ via the EM-test we are

Table 3.10: BIC for the corresponding models for series 1947(1)–1984(1).

BIC	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
$p=1$	**455.78**	476.21	468.33	469.93
$p=2$	460.13	477.64	472.95	474.55
$p=3$	463.18	486.29	474.72	477.49
$p=4$	465.50	492.24	476.56	479.00

not able to reject the hypothesis of no regime switch in the intercept.
Considering the subseries from 1984(2)–2002(3), the BIC favors a purely linear autoregressive model of order $p = 2$, see Table 3.12 whereas the AIC favors model $\mathcal{M}_2$ with order $p = 2$, directly followed by model $\mathcal{M}_4$ with order $p = 2$, see Table 3.13.

Therefore, we test for homogeneity in model $\mathcal{M}_4$ by the EM-test. Since the sample size is rather small we cannot show definitive evidence of a regime switch in the intercept in model $\mathcal{M}_4$ for this subseries. Fitted values and results of the EM-test are shown in Table 3.14.

Table 3.11: AIC for the corresponding models for series 1947(1)–1984(1).

AIC	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
$p = 1$	**446.87**	452.45	450.52	452.11
$p = 2$	448.25	447.94	452.16	453.76
$p = 3$	448.33	450.65	450.96	453.74
$p = 4$	447.68	450.66	449.83	452.27

Table 3.12: BIC for the corresponding models for series 1984(2)–2002(3).

BIC	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
$p = 1$	120.62	128.19	133.53	124.23
$p = 2$	**119.80**	132.36	132.72	125.78
$p = 3$	123.74	140.45	136.66	130.09
$p = 4$	128.05	147.22	140.96	133.91

Table 3.13: AIC for the corresponding models for series 1984(2)–2002(3).

AIC	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
$p = 1$	113.71	109.75	119.71	110.40
$p = 2$	110.59	**109.32**	116.59	109.66
$p = 3$	112.22	112.80	118.22	111.65
$p = 4$	114.22	114.97	120.22	113.17

Table 3.14: Fits of mixture AR(p) models with possibly switching intercept using modified quasi MLE with normal innvoations and results of the EM-test (subseries 1984(2)–2002(3))

	$\overline{\zeta}_1$	$\overline{\zeta}_2$	$\overline{\phi}_1$	$\overline{\phi}_2$	$\overline{\phi}_3$	$\overline{\phi}_4$	$\overline{\sigma}$	$EM_n^{(2)}$	p-value
$p = 1$	0.06	0.82	0.41				0.33	3.12	0.08
$p = 2$	-0.06	0.71	0.28	0.26			0.30	4.41	0.04
$p = 3$	-0.07	0.70	0.28	0.26	0.02		0.30	4.09	0.04
$p = 4$	-0.10	0.68	0.28	0.24	0.01	0.05	0.30	4.37	0.04

3.6 Proofs

To prove Theorem 3.1 we need some additional lemmas. Since we assume that the innovations $(\epsilon_k)_k$ are independent and identically normally distributed with expectation 0 and scale parameter σ, the conditional density (w.r.t. Lebesgue measure on $\mathbb{R}$) of X_t given $X_{t-1}^p = x_{t-1}^p$ and $S_t = i$ is

$$g(x_t|x_{t-1}^p;\zeta_i,\boldsymbol{\phi},\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x_t-\zeta_i-\sum_{j=1}^p \phi_j x_{t-j})^2}{2\sigma^2}\right).$$

In the following, let $(\overline{\alpha},\overline{\zeta}_1,\overline{\zeta}_2,\overline{\boldsymbol{\phi}},\overline{\sigma})$ be any EM-estimator. Then we write that a statement holds for example for $\overline{\alpha}$ if and only if it holds for every $\alpha_j^{(k)}$, $j=1,\dots,J$ and $k=1,\dots,K$.

Lemma 3.1. *For each given $\overline{\alpha}\in(0,0.5]$ we have under the null model*

$$\begin{aligned}\overline{\sigma}-\sigma_0 &= o_P(1), & \overline{\boldsymbol{\phi}}-\boldsymbol{\phi}_0 &= o_P(1),\\ \overline{\zeta}_1-\zeta_0 &= o_P(1), & \overline{\zeta}_2-\zeta_0 &= o_P(1).\end{aligned}$$

Proof. Since we assume $(X_k)_k$ to be a causal AR(p) process under the null model we know that the order of the autoregressive process is uniquely defined and that the parameters are identifiable (cf. Kreiss and Neuhaus, 2006). Assuming $\sigma_0\in[\delta,\infty)$, $\delta>0$, we have

$$\begin{aligned}&E\left|\log\big(g(X_1|X_0^p;\zeta_0,\boldsymbol{\phi}_0,\sigma_0)\big)\right|\\ =\;& E\left[\left|\log\left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\exp\left(-\frac{1}{2}\frac{(X_1-\zeta_0-\sum_{j=1}^p\phi_{j,0}X_{1-j})^2}{\sigma_0^2}\right)\right)\right|\right]\\ \le\;& E\left[\frac{1}{2}\frac{(X_1-\zeta_0-\sum_{j=1}^p\phi_{j,0}X_{1-j})^2}{\sigma_0^2}\right]+\left|\frac{1}{2}\log(2\pi\sigma_0^2)\right|\\ =\;& 1/2+\left|\frac{1}{2}\log(2\pi\sigma_0^2)\right|<\infty.\end{aligned}$$

Therefore $(\log g(X_t|X_{t-1}^p;\zeta_0,\boldsymbol{\phi}_0,\sigma_0))_t$ obeys the strong law of large numbers by the ergodic theorem (cf. Krengel, 1985). Further, we have

$$l_n(\overline{\alpha},\overline{\zeta}_1,\overline{\zeta}_2,\overline{\boldsymbol{\phi}},\overline{\sigma})-l_n(0.5,\zeta_0,\zeta_0,\boldsymbol{\phi}_0,\sigma_0)\ge\min_{j=1,\dots,J}\{p(\alpha_j)-p(0.5)\}>-\infty \tag{3.6.1}$$

since $0.5 = \arg\max_\alpha p(\alpha)$ and

$$\begin{aligned} l_n(\alpha_j^{(k)}, \zeta_{1j}^{(k)}, \zeta_{2j}^{(k)}, \boldsymbol{\phi}_j^{(k)}, \sigma_j^{(k)}) + p(\alpha_j^{(k)}) &= pl_n(\alpha_j^{(k)}, \zeta_{1j}^{(k)}, \zeta_{2j}^{(k)}, \boldsymbol{\phi}_j^{(k)}, \sigma_j^{(k)}) \\ &\geq pl_n(\alpha_j^{(0)}, \zeta_{1j}^{(0)}, \zeta_{2j}^{(0)}, \boldsymbol{\phi}_j^{(0)}, \sigma_j^{(0)}) \\ &\geq pl_n(\alpha_j, \zeta_0, \zeta_0, \boldsymbol{\phi}_0, \sigma_0) \\ &= l_n(\alpha_j, \zeta_0, \zeta_0, \boldsymbol{\phi}_0, \sigma_0) + p(\alpha_j) \end{aligned}$$

for every $j = 1, \ldots, J$ and $k = 1, \ldots, K$ by the *EM-property*. Now the result follows using the argument as in Theorem 2 for the i.i.d. case in Wald (1949). □

From now on we assume without loss of generality $\zeta_0 = 0$ and $\sigma_0 = 1$. Let

$$\begin{aligned} Y_t &:= \left. \frac{\frac{\partial}{\partial\zeta} g(X_t|X_{t-1}^p; \zeta, \boldsymbol{\phi}_0, 1)}{g(X_t|X_{t-1}^p; 0, \boldsymbol{\phi}_0, 1)} \right|_{\zeta=0}, \\ Z_t &:= \frac{1}{2} \left. \frac{\frac{\partial^2}{\partial^2\zeta} g(X_t|X_{t-1}^p; \zeta, \boldsymbol{\phi}_0, 1)}{g(X_t|X_{t-1}^p; 0, \boldsymbol{\phi}_0, 1)} \right|_{\zeta=0}, \\ U_t &:= \frac{1}{6} \left. \frac{\frac{\partial^3}{\partial^3\zeta} g(X_t|X_{t-1}^p; \zeta, \boldsymbol{\phi}_0, 1)}{g(X_t|X_{t-1}^p; 0, \boldsymbol{\phi}_0, 1)} \right|_{\zeta=0}, \\ V_t &:= \frac{1}{24} \left. \frac{\frac{\partial^4}{\partial^4\zeta} g(X_t|X_{t-1}^p; \zeta, \boldsymbol{\phi}_0, 1)}{g(X_t|X_{t-1}^p; 0, \boldsymbol{\phi}_0, 1)} \right|_{\zeta=0} \end{aligned}$$

and

$$\begin{aligned} W_{\tau t} &:= \frac{\partial_{\phi_\tau} g(X_t|X_{t-1}^p; 0, \boldsymbol{\phi}_0, 1)}{g(X_t|X_{t-1}^p; 0, \boldsymbol{\phi}_0, 1)} \\ &= X_{t-\tau} Y_t, \quad \tau = 1, \ldots, p, \end{aligned}$$

writing for fixed $(x, y^p) \in \mathbb{R} \times \mathbb{R}^p$ $\partial_{\phi_\tau} g(x|y^p; 0, \boldsymbol{\phi}_0, 1)$ for the partial derivative of the function $g(x|y^p; 0, \cdot, 1) : \boldsymbol{H} \to \mathbb{R}_{\geq 0}$ with respect to the argument ϕ_τ, $\tau = 1, \ldots, p$, evaluated at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$.

Computing these derivatives yields

$$\begin{aligned} Y_t &= \epsilon_t, \\ Z_t &= (\epsilon_t^2 - 1)/2, \\ U_t &= (\epsilon_t^3 - 3\epsilon_t)/6, \\ V_t &= (\epsilon_t^4 - 6\epsilon_t^2 + 3)/24, \\ W_{\tau t} &= X_{t-\tau}\epsilon_t, \quad \tau = 1, \ldots, p. \end{aligned}$$

To ensure readability we restrict our attention to the case $p = 1$. If there will be more than notational change for $p > 1$ then we will note this explicitly.

Lemma 3.2. *For each $\alpha_j \in (0, 0.5]$ we have under the null model*

$$\overline{\sigma}^2 - 1 = O_P(n^{-1/4}), \qquad \overline{\phi} - \phi_0 = O_P(n^{-1/2}),$$
$$\overline{\zeta}_1 = O_P(n^{-1/8}), \qquad \overline{\zeta}_2 = O_P(n^{-1/8}),$$

whenever $\overline{\alpha} - \alpha_j = o_P(1)$.

Proof. Let $\overline{m}_k = (1-\overline{\alpha})\overline{\zeta}_1^k + \overline{\alpha}\overline{\zeta}_2^k$, $k = 1, \ldots, 6$ and

$$\begin{aligned}
\overline{s}_1 &= \overline{m}_1, \\
\overline{s}_2 &= \overline{m}_2 + (\overline{\sigma}^2 - 1), \\
\overline{s}_3 &= \overline{m}_3, \\
\overline{s}_4 &= \overline{m}_4 - 3\overline{m}_2^2, \\
\overline{s}_5 &= (\overline{\phi} - \phi_0).
\end{aligned}$$

Note here that we absorbed the term $3(\overline{\sigma}^2 - 1)\overline{m}_1$ (which appears in the Taylor expansion as coefficient of U_t) into the remainder (3.6.4) as in Chen and Chen (2003), since by CLT and Lemma 3.1, $\sum_{t=1}^n 3(\overline{\sigma}^2 - 1)\overline{m}_1 U_t = (\overline{\sigma}^2 - 1)\overline{m}_1 O_P(n^{1/2}) = o_P(n^{1/2}\overline{m}_1)$. Due to the inequality $|x| \leq 1 + x^2$, we have $o_P(n^{1/2}\overline{m}_1) = o_P(1) + n o_P(\overline{m}_1^2)$ which is part of the remainder (3.6.4). The coefficient $\overline{s}_4$ of V_t would be $3(\overline{\sigma}^2 - 1)^2 + \overline{m}_4 + 6(\overline{\sigma}^2 - 1)\overline{m}_2$ rather than $\overline{m}_4 - 3\overline{m}_2^2$. Simple algebra gives

$$3(\overline{\sigma}^2 - 1)^2 + \overline{m}_4 + 6(\overline{\sigma}^2 - 1)\overline{m}_2 = 3\underbrace{(\overline{\sigma}^2 - 1 + \overline{m}_2)^2}_{=\overline{s}_2^2} + \overline{m}_4 - 3\overline{m}_2^2.$$

Using the CLT we have $3\sum_{t=1}^n \overline{s}_2^2 V_t = n^{1/2}\overline{s}_2^2 O_P(1)$, being just the first term of (3.6.4). Our aim is to find an asymptotic upper bound for

$$2\{pl_n(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}) - pl_n(0.5, 0, 0, \phi_0, 1)\}.$$

To this end, we write $2\{pl_n(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}) - pl_n(0.5, 0, 0, \phi_0, 1)\} = 2\sum_{t=1}^n \log(1 + \overline{\delta}_t) + 2\{p(\overline{\alpha}) - p(0.5)\}$, with

$$\overline{\delta}_t = (1-\overline{\alpha})\left\{\frac{g(X_t|X_{t-1}; \overline{\zeta}_1, \overline{\phi}, \overline{\sigma})}{g(X_t|X_{t-1}; 0, \phi_0, 1)} - 1\right\} + \overline{\alpha}\left\{\frac{g(X_t|X_{t-1}; \overline{\zeta}_2, \overline{\phi}, \overline{\sigma})}{g(X_t|X_{t-1}; 0, \phi_0, 1)} - 1\right\}.$$

Using Taylor expansion we get

$$\overline{\delta}_t = \overline{s}_1 Y_t + \overline{s}_2 Z_t + \overline{s}_3 U_t + \overline{s}_4 V_t + \overline{s}_5 W_{1t} + \overline{\epsilon}_{tn} \tag{3.6.2}$$

with an appropriate remainder $\bar{\epsilon}_{tn}$. Due to Lemma 3.1 we can assume that $(\bar{\zeta}_1, \bar{\zeta}_2, \bar{\phi}, \bar{\sigma})$ is in a small neighborhood of $(0, 0, \phi_0, 1)$. Further, note that the remainders resulting from the square and cubic sums in (3.6.3) will be of the same or higher order than the remainder $\bar{\epsilon}_n = \sum_{t=1}^n \bar{\epsilon}_{tn}$ from the linear sum and can thus be omitted. Also, the other terms (including the cross-product terms) in the expansion (3.6.2) will be absorbed in the remainder term $\bar{\epsilon}_n$ since by Lemma 3.1, the CLT for stationary and ergodic martingale differences and the inequality $|x| \leq 1 + x^2$, e.g.

$$\begin{aligned}
O_P(n^{1/2})(\bar{\sigma}^2 - 1)\bar{m}_1 &= o_P(n^{1/2})\bar{m}_1 = o_P(1) + n\bar{m}_1^2 o_P(1), \\
O_P(n^{1/2})(\bar{\sigma}^2 - 1)(\bar{\phi} - \phi_0) &= o_P(n^{1/2})(\bar{\phi} - \phi_0) = o_P(1) + n(\bar{\phi} - \phi_0)^2 o_P(1), \\
O_P(n^{1/2})\bar{m}_2 &= o_P(n^{1/2})\bar{m}_1 = o_P(1) + n\bar{m}_1^2 o_P(1), \\
O_P(n^{1/2})\bar{m}_4 &= o_P(n^{1/2})\bar{m}_3 = o_P(1) + n\bar{m}_3^2 o_P(1).
\end{aligned}$$

Using the inequality $\log(1 + x) \leq x - \frac{1}{2}x^2 + \frac{1}{3}x^3$ we get

$$\begin{aligned}
&2\{pl_n(\bar{\alpha}, \bar{\zeta}_1, \bar{\zeta}_2, \bar{\phi}, \bar{\sigma}) - pl_n(0.5, 0, 0, \phi_0, 1)\} \\
&\leq 2\sum_{t=1}^n \{\bar{s}_1 Y_t + \bar{s}_2 Z_t + \bar{s}_3 U_t + \bar{s}_4 V_t + \bar{s}_5 W_{1t}\} \\
&\quad - \sum_{t=1}^n \{\bar{s}_1 Y_t + \bar{s}_2 Z_t + \bar{s}_3 U_t + \bar{s}_4 V_t + \bar{s}_5 W_{1t}\}^2 \\
&\quad + \frac{2}{3}\sum_{t=1}^n \{\bar{s}_1 Y_t + \bar{s}_2 Z_t + \bar{s}_3 U_t + \bar{s}_4 V_t + \bar{s}_5 W_{1t}\}^3 + \bar{\epsilon}_n \\
&\quad + 2\{p(\bar{\alpha}) - p(0.5)\}
\end{aligned} \tag{3.6.3}$$

with remainder $\bar{\epsilon}_n = \sum_{t=1}^n \bar{\epsilon}_{tn}$ satisfying

$$\begin{aligned}
\bar{\epsilon}_n = &\, n^{1/2}\bar{s}_2^2 O_P(1) + n^{1/2}(\bar{\sigma}^2 - 1)^3 O_P(1) + n(\bar{\phi} - \phi_0)^2 o_P(1) \\
&+ n(\bar{m}_1^2 + \bar{m}_3^2) o_P(1) + n^{1/2}(|\bar{m}_5| + \bar{m}_6) O_P(1) + o_P(1).
\end{aligned} \tag{3.6.4}$$

We use the following lemma (together with Lemma 3.1) to show that

$$\bar{\epsilon}_n = o_P(1) + n o_P\Big(\sum_{j=1}^5 \bar{s}_j^2\Big)$$

and therefore (together with the fact that the covariance matrix of $(Y_t, Z_t, U_t, V_t, W_{1t})$ is non-degenerate) the remainder $\bar{\epsilon}_n$ can be absorbed by the quadratic term in (3.6.3).

Lemma 3.3. *Let* $(\bar{\alpha}, \bar{\zeta}_1, \bar{\zeta}_2, \bar{\phi}, \bar{\sigma})$ *be an EM-estimator of* $(\alpha, \zeta_1, \zeta_2, \phi, \sigma)$ *and* $\bar{\alpha} \in [\delta', 1 - \delta']$

for some $\delta' \in (0, 0.5]$. *Then*

$$\overline{\zeta}_1^4 = O_P\Big(\sum_{j=1}^{5} |\overline{s}_j|\Big),\ \overline{\zeta}_2^4 = O_P\Big(\sum_{j=1}^{5} |\overline{s}_j|\Big) \text{ and } (\overline{\sigma}^2 - 1)^2 = O_P\Big(\sum_{j=1}^{5} |\overline{s}_j|\Big)$$

holds under the null model.

Proof. By the definition of $\overline{s}_1$, we have

$$\overline{\zeta}_1 = \frac{1}{1-\overline{\alpha}}\overline{s}_1 - \frac{\overline{\alpha}}{1-\overline{\alpha}}\overline{\zeta}_2. \tag{3.6.5}$$

Plugging (3.6.5) in the definition of $\overline{s}_4$, we get

$$\begin{aligned}\overline{s}_4 = & \frac{\overline{\alpha}(1-6\overline{\alpha}+6\overline{\alpha}^2)}{(1-\overline{\alpha}^3)}\overline{\zeta}_2^4 \\ & + \frac{\big(16\overline{\alpha}^3\overline{\zeta}_2^3 - 12\overline{s}_1\overline{\alpha}^3\overline{\zeta}_2^2 - 12\overline{\alpha}^2\overline{\zeta}_2^3 + 12\overline{s}_1^2\overline{\alpha}^2\overline{\zeta}_2\big)}{(\overline{\alpha}-1)^3}\overline{s}_1 \\ & + \frac{\big(6\overline{s}_1\overline{\alpha}\overline{\zeta}_2^2 - 8\overline{s}_1^2\overline{\alpha}\overline{\zeta}_2 - 3\overline{\alpha}\overline{s}_1^3 + 2\overline{s}_1^3\big)}{(\overline{\alpha}-1)^3}\overline{s}_1\end{aligned}$$

which leads to

$$\overline{s}_4 = \frac{\overline{\alpha}(1-6\overline{\alpha}+6\overline{\alpha}^2)}{(1-\overline{\alpha})^3}\overline{\zeta}_2^4 + o_P(|\overline{s}_1|) \tag{3.6.6}$$

due to Lemma 3.1. This gives $\overline{\zeta}_2^4 = O_P\Big(\sum_{j=1}^5 |\overline{s}_j|\Big)$ since $\overline{\alpha}$ is assumed to be in $[\delta', 1-\delta']$. By (3.6.5) we also conclude $\overline{\zeta}_1^4 = O_P\Big(\sum_{j=1}^5 |\overline{s}_j|\Big)$. Applying the inequality $(a+b)^2 \le 2(a^2+b^2)$ for $a, b \ge 0$ and Lemma 3.1, we obtain

$$\begin{aligned} 0 \le (\overline{\sigma}^2-1)^2 &= (\overline{s}_2 - \overline{m}_2)^2 \le (|\overline{s}_2| + \overline{m}_2)^2 \\ &\le 2|\overline{s}_2|^2 + 2\overline{m}_2^2 = o_P(|\overline{s}_2|) + 2\big((1-\overline{\alpha})\overline{\zeta}_1^2 + \overline{\alpha}\overline{\zeta}_2^2\big)^2 \\ &\le o_P(|\overline{s}_2|) + 4(1-\overline{\alpha})^2\overline{\zeta}_1^4 + 4\overline{\alpha}^2\overline{\zeta}_2^4 \\ &= O_P\Big(\sum_{j=1}^{5} |\overline{s}_j|\Big), \end{aligned}$$

where the last equality follows by $\overline{\zeta}_i^4 = O_P\Big(\sum_{j=1}^5 |\overline{s}_j|\Big)$, $i = 1, 2$, and the assumption $\overline{\alpha} \in [\delta', 1-\delta']$. Therefore, we have $(\overline{\sigma}^2-1)^2 = O_P\Big(\sum_{j=1}^5 |\overline{s}_j|\Big)$. □

In the following we show how to use this lemma to prove that the remainder $\overline{\epsilon}_n$ can be absorbed by the quadratic term in (3.6.3). Regarding the particular summands in (3.6.4) we see

$$
\begin{aligned}
n^{1/2}\overline{s}_2^2 O_P(1) &= n\overline{s}_2^2 o_P(1),\\
n^{1/2}(\overline{\sigma}^2-1)^3 O_P(1) &= n^{1/2}(\overline{\sigma}^2-1)^2 o_P(1) = n^{1/2} o_P\big(\sum_{j=1}^{5}|\overline{s}_j|\big)\\
&= o_P(1) + n o_P\big(\sum_{j=1}^{5}\overline{s}_j^2\big),\\
n(\overline{\phi}-\phi_0)^2 o_P(1) &= n\overline{s}_5^2 o_P(1),\\
n(\overline{m}_1^2+\overline{m}_3^2) o_P(1) &= n(\overline{s}_1^2+\overline{s}_3^2) o_P(1),\\
n^{1/2}|\overline{m}_5| O_P(1) &= n^{1/2}\overline{m}_4 o_P(1) = n^{1/2} o_P\big(\sum_{j=1}^{5}|\overline{s}_j|\big)\\
&= o_P(1) + n o_P\big(\sum_{j=1}^{5}\overline{s}_j^2\big),\\
n^{1/2}\overline{m}_6 O_P(1) &= n^{1/2}\overline{m}_4 o_P(1) = n^{1/2} o_P\big(\sum_{j=1}^{5}|\overline{s}_j|\big)\\
&= o_P(1) + n o_P\big(\sum_{j=1}^{5}\overline{s}_j^2\big),
\end{aligned}
$$

applying the inequality $|x| \le 1 + x^2$, Lemma 3.1 and Lemma 3.3.

In the following, we show that the cubic term is negligible compared to the quadratic term in the expansion (3.6.3). Using the inequality $(a+b)^3 \le 4(a^3+b^3)$ for non-negative a, b repeatedly, ergodic theorem and Lemma 3.1, we get

$$
\begin{aligned}
& \Big|\{\overline{s}_1\sum_{t=1}^{n} Y_t + \overline{s}_2\sum_{t=1}^{n} Z_t + \overline{s}_3\sum_{t=1}^{n} U_t + \overline{s}_4\sum_{t=1}^{n} V_t + \overline{s}_5\sum_{t=1}^{n} W_{1t}\}^3\Big|\\
\le\ & \mathrm{const}\{|\overline{s}_1|^3\sum_{t=1}^{n}|Y_t|^3 + |\overline{s}_2|^3\sum_{t=1}^{n}|Z_t|^3 + |\overline{s}_3|^3\sum_{t=1}^{n}|U_t|^3\\
& + |\overline{s}_4|^3\sum_{t=1}^{n}|V_t|^3 + |\overline{s}_5|^3\sum_{t=1}^{n}|W_{1t}|^3\}\\
=\ & O_P(n)\{\sum_{j=1}^{5}|\overline{s}_j|^3\}
\end{aligned}
$$

$$= o_P(n)\{\sum_{j=1}^{5} |\overline{s}_j|^2\}.$$

Since the covariance matrix of $(Y_t, Z_t, U_t, V_t, W_{1t})$ is non-degenerate, the right-hand side of inequation (3.6.3) reduces to

$$\begin{aligned}
&2\{pl_n(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}) - pl_n(0.5, 0, 0, \phi_0, 1)\}\\
&\leq 2\sum_{t=1}^{n}\{\overline{s}_1 Y_t + \overline{s}_2 Z_t + \overline{s}_3 U_t + \overline{s}_4 V_t + \overline{s}_5 W_{1t}\}\\
&\quad - \sum_{t=1}^{n}\{\overline{s}_1 Y_t + \overline{s}_2 Z_t + \overline{s}_3 U_t + \overline{s}_4 V_t + \overline{s}_5 W_{1t}\}^2\{1 + o_P(1)\}\\
&\quad + 2\{p(\overline{\alpha}) - p(0.5)\} + o_P(1).
\end{aligned} \tag{3.6.7}$$

Note that Y_t, Z_t, U_t, V_t and W_{1t} are mutually orthogonal (see Section 3.6.1). Therefore

$$\begin{aligned}
&2\{pl_n(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}) - pl_n(0.5, 0, 0, \phi_0, 1)\}\\
&= 2\{l_n(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}) - l_n(0.5, 0, 0, \phi_0, 1)\} + 2\{p(\overline{\alpha}) + p(0.5)\}\\
&\leq 2\overline{s}_1 \sum_{t=1}^{n} Y_t - \overline{s}_1^2 \sum_{t=1}^{n} Y_t^2\{1 + o_p(1)\}\\
&\quad + 2\overline{s}_2 \sum_{t=1}^{n} Z_t - \overline{s}_2^2 \sum_{t=1}^{n} Z_t^2\{1 + o_p(1)\}\\
&\quad + 2\overline{s}_3 \sum_{t=1}^{n} U_t - \overline{s}_3^2 \sum_{t=1}^{n} U_t^2\{1 + o_p(1)\}\\
&\quad + 2\overline{s}_4 \sum_{t=1}^{n} V_t - \overline{s}_4^2 \sum_{t=1}^{n} V_t^2\{1 + o_p(1)\}\\
&\quad + 2\overline{s}_5 \sum_{t=1}^{n} W_{1t} - \overline{s}_5^2 \sum_{t=1}^{n} W_{1t}^2\{1 + o_p(1)\}\\
&\quad + 2\{p(\alpha_j) + p(0.5)\} + o_P(1).
\end{aligned} \tag{3.6.8}$$

The last step is due to the assumption $\overline{\alpha} - \alpha_j = o_P(1)$ and the fact that $p(\alpha)$ is a continuous function in α.

Note that

$$2\overline{s}_1 \sum_{t=1}^{n} Y_t - \overline{s}_1^2 \sum_{t=1}^{n} Y_t^2\{1 + o_P(1)\} \leq \frac{(\sum_{t=1}^{n} Y_t)^2}{\sum_{t=1}^{n} Y_t^2}\{1 + o_P(1)\} = O_P(1).$$

Thereby, the last equality follows from the CLT (applied to the numerator) and the SLLN (applied to the denominator).
The same argumentation holds true if we replace Y_t by Z_t, U_t or V_t. Using the CLT for stationary and ergodic martingale differences and ergodic theorem we get

$$2\overline{s}_5 \sum_{t=1}^{n} W_{\tau t} - \overline{s}_5^2 \sum_{t=1}^{n} W_{\tau t}^2 \{1 + o_P(1)\} \leq \frac{(\sum_{t=1}^{n} W_{\tau t})^2}{\sum_{t=1}^{n} W_{\tau t}^2} \{1 + o_P(1)\} = O_P(1).$$

By the *EM-property*

$$\begin{aligned} 0 &\leq 2\{pl_n(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}) - pl_n(0.5, 0, 0, \phi_0, 1)\} \\ &\leq 2\overline{s}_1 \sum_{t=1}^{n} Y_t - \overline{s}_1^2 \sum_{t=1}^{n} Y_t^2 \{1 + o_P(1)\} + O_P(1) \\ &= O_P(1) \end{aligned}$$

and so

$$2\overline{s}_1 \sum_{t=1}^{n} Y_t - \overline{s}_1^2 \sum_{t=1}^{n} Y_t^2 \{1 + o_P(1)\} = O_P(1)$$

which leads to $\overline{s}_1 = O_P(n^{-1/2})$. Analogously, we get

$$\overline{s}_j = O_P(n^{-1/2}), \quad j = 2, 3, 4, 5. \tag{3.6.9}$$

By the definition of $\overline{s}_5$ it immediately follows that

$$\overline{\phi} - \phi = O_P(n^{-1/2})$$

while due to the assumption $\overline{\alpha} - \alpha_j = o_P(1)$ and Lemma 3.3 we get

$$\overline{\sigma}^2 - 1 = O_P(n^{-1/4}), \quad \overline{\zeta}_i = O_P(n^{-1/8}), \, i = 1, 2.$$

$\square$

Remark 3.2. If the order of the AR(p) process under the null model is higher than 1 the right-hand side of (3.6.7) will be replaced by

$$\begin{aligned} &2 \sum_{t=1}^{n} \{\overline{s}_1 Y_t + \overline{s}_2 Z_t + \overline{s}_3 U_t + \overline{s}_4 V_t + \big(\sum_{\tau=1}^{p} \overline{s}_{4+\tau} W_{\tau t}\big)\} \\ &- \sum_{t=1}^{n} \{\overline{s}_1 Y_t + \overline{s}_2 Z_t + \overline{s}_3 U_t + \overline{s}_4 V_t + \big(\sum_{\tau=1}^{p} \overline{s}_{4+\tau} W_{\tau t}\big)\}^2 \{1 + o_P(1)\} \\ &+ 2\{p(\overline{\alpha}) - p(0.5)\} + o_P(1) \end{aligned} \tag{3.6.10}$$

with $\overline{s}_{4+\tau} = \overline{\phi}_\tau - \phi_{\tau,0}$, $\tau = 1, \ldots, p$. We orthogonalize $W_{1t}, \ldots, W_{pt}$ (using Gram-Schmidt), such that

$$\begin{aligned} \tilde{W}_{1t} &= W_{1t}, \quad \tilde{W}_{2t} = W_{2t} - \frac{E\tilde{W}_{1t}W_{2t}}{E\tilde{W}_{1t}^2}\tilde{W}_{1t}, \ldots, \\ \tilde{W}_{pt} &= W_{pt} - \sum_{\tau=1}^{p-1} \frac{E\tilde{W}_{\tau t}W_{pt}}{E\tilde{W}_{\tau t}^2}\tilde{W}_{\tau t} \end{aligned} \tag{3.6.11}$$

to obtain an analogous expansion to (3.6.8). This orthogonalization does not affect the random variables Y_t, V_t, Z_t and U_t since for each given $\tau = 1, \ldots, p$, these and $W_{\tau t}$ are already mutually orthogonal (see Section 3.6.1). Therefore (3.6.10) can be written as

$$\begin{aligned} &2\sum_{t=1}^{n}\{\overline{s}_1 Y_t + \overline{s}_2 Z_t + \overline{s}_3 U_t + \overline{s}_4 V_t + \big(\sum_{\tau=1}^{p} \overline{s}'_{4+\tau}\tilde{W}_{\tau t}\big)\} \\ &- \sum_{t=1}^{n}\{\overline{s}_1 Y_t + \overline{s}_2 Z_t + \overline{s}_3 U_t + \overline{s}_4 V_t + \big(\sum_{\tau=1}^{p} \overline{s}'_{4+\tau}\tilde{W}_{\tau t}\big)\}^2\{1 + o_P(1)\} \\ &+ 2\{p(\overline{\alpha}) - p(0.5)\} + o_P(1), \end{aligned} \tag{3.6.12}$$

where $\overline{s}'_{4+\tau}$ is a linear combination of $\overline{s}_{4+\tau}, \ldots, \overline{s}_{4+p}$, particularly $\overline{s}'_{4+p} = \overline{s}_{4+p} = \overline{\phi}_p - \phi_{p,0}$. Hence, the right-hand side of (3.6.8) will be

$$\begin{aligned} &2\overline{s}_1 \sum_{t=1}^{n} Y_t - \overline{s}_1^2 \sum_{t=1}^{n} Y_t^2\{1 + o_p(1)\} \\ &+ 2\overline{s}_2 \sum_{t=1}^{n} Z_t - \overline{s}_2^2 \sum_{t=1}^{n} Z_t^2\{1 + o_p(1)\} \\ &+ 2\overline{s}_3 \sum_{t=1}^{n} U_t - \overline{s}_3^2 \sum_{t=1}^{n} U_t^2\{1 + o_p(1)\} \\ &+ 2\overline{s}_4 \sum_{t=1}^{n} V_t - \overline{s}_4^2 \sum_{t=1}^{n} V_t^2\{1 + o_p(1)\} \\ &+ 2\overline{s}'_5 \sum_{t=1}^{n} \tilde{W}_{1t} - \overline{s}'^2_5 \sum_{t=1}^{n} \tilde{W}_{1t}^2\{1 + o_p(1)\} \\ &\vdots \\ &+ 2\overline{s}'_{4+p} \sum_{t=1}^{n} \tilde{W}_{1t} - \overline{s}'^2_{4+p} \sum_{t=1}^{n} \tilde{W}_{pt}^2\{1 + o_p(1)\} \\ &+ 2\{p(\alpha_j) + p(0.5)\} + o_P(1) \end{aligned} \tag{3.6.13}$$

for $p > 1$.

Let $(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma})$ be some EM-estimator. We define

$$\begin{aligned} H_n(\alpha) &= \left(n - \sum_{t=1}^{n} \overline{w}_t\right) \log(1-\alpha) + \sum_{t=1}^{n} \overline{w}_t \log(\alpha) + p(\alpha) \\ &=: R_n(\alpha) + p(\alpha), \end{aligned}$$

where

$$\overline{w}_t = \frac{\overline{\alpha} g(X_t|X_{t-1}; \overline{\zeta}_2, \overline{\phi}, \overline{\sigma})}{(1-\overline{\alpha}) g(X_t|X_{t-1}; \overline{\zeta}_1, \overline{\phi}, \overline{\sigma}) + \overline{\alpha} g(X_t|X_{t-1}; \overline{\zeta}_2, \overline{\phi}, \overline{\sigma})}.$$

Let $\overline{\alpha}^* = \arg\max_{\alpha \in [0,1]} H_n(\alpha)$. The following lemma shows that if $\overline{\alpha} - \alpha_j = O_P(n^{-1/4})$ holds true for any estimator $\overline{\alpha}$ then also for the estimator $\overline{\alpha}^*$ maximizing $H_n(\alpha)$.

Lemma 3.4. *Let $(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma})$ be an EM-estimator. If $\overline{\alpha} - \alpha_j = O_P(n^{-1/4})$ for some $\alpha_j \in (0,1)$, then under the null model, we have*

$$\overline{\alpha}^* - \alpha_j = O_P(n^{-1/4}).$$

Proof. Let $\widehat{\alpha} = (1/n) \sum_{t=1}^{n} \overline{w}_t$ be the maximizer of $R_n(\alpha)$. We have

$$|\widehat{\alpha} - \overline{\alpha}| = \frac{\overline{\alpha}(1-\overline{\alpha})}{n} \left| \sum_{t=1}^{n} \frac{g(X_t|X_{t-1}; \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}) - g(X_t|X_{t-1}; \overline{\zeta}_1, \overline{\phi}, \overline{\sigma})}{(1-\overline{\alpha}) g(X_t|X_{t-1}; \overline{\zeta}_1, \overline{\phi}, \overline{\sigma}) + \overline{\alpha} g(X_t|X_{t-1}; \overline{\zeta}_2, \overline{\phi}, \overline{\sigma})} \right|.$$

Note that $(\overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma})$ is in a small neighborhood of $(0, 0, \phi_0, 1)$ by Lemma 3.1. Defining

$$\tilde{\gamma}(\overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}) = \sum_{t=1}^{n} \frac{g(X_t|X_{t-1}; \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}) - g(X_t|X_{t-1}; \overline{\zeta}_1, \overline{\phi}, \overline{\sigma})}{(1-\overline{\alpha}) g(X_t|X_{t-1}; \overline{\zeta}_1, \overline{\phi}, \overline{\sigma}) + \overline{\alpha} g(X_t|X_{t-1}; \overline{\zeta}_2, \overline{\phi}, \overline{\sigma})}$$

and expanding $\tilde{\gamma}(\cdot)$ at $(0, 0, \phi_0, 1)$ to order 1 using SLLN, CLT and the order information in Lemma 3.2, we get

$$\begin{aligned} |\widehat{\alpha} - \overline{\alpha}| &= \frac{\overline{\alpha}(1-\overline{\alpha})}{n} \left| (\overline{\zeta}_1 - \overline{\zeta}_2) \sum_{t=1}^{n} Y_t + O_p(n) \{ \overline{\zeta}_1^2 + \overline{\zeta}_2^2 \} \right| \\ &= \frac{\overline{\alpha}(1-\overline{\alpha})}{n} \left| (\overline{\zeta}_1 - \overline{\zeta}_2) \sum_{t=1}^{n} Y_t + O_p(n^{3/4}) \right| \\ &= O_P(n^{-1/4}). \end{aligned} \tag{3.6.14}$$

By the triangular inequality it suffices to prove $\overline{\alpha}^* - \widehat{\alpha} = O_P(n^{-1/4})$ since we have $\overline{\alpha} - \alpha_j = O_P(n^{-1/4})$ by assumption and $\widehat{\alpha} - \overline{\alpha} = O_P(n^{-1/4})$.

Our next step is to show by contradiction that

$$\overline{\alpha}^* - \widehat{\alpha} = o_P(1). \tag{3.6.15}$$

Note that by assumption and (3.6.14)

$$\overline{\alpha} - \alpha_j = O_P(n^{-1/4}) = o_P(1), \quad \widehat{\alpha} - \overline{\alpha} = O_P(n^{-1/4}) = o_P(1), \tag{3.6.16}$$

and that $R_n(\alpha)$ is a binomial log-likelihood which attains its maximum at $\widehat{\alpha}$.

First we assume $\overline{\alpha}^* \geq \widehat{\alpha} + 2\epsilon$ for any $\epsilon > 0$. Since $R_n(\alpha)$ is a decreasing function in α (for $\alpha \geq \widehat{\alpha}$) and attains its maximum at $\widehat{\alpha}$, we have for $\alpha \geq \widehat{\alpha} + 2\epsilon$

$$R_n(\alpha) - R_n(\widehat{\alpha}) \leq R_n(\widehat{\alpha} + 2\epsilon) - R_n(\widehat{\alpha} + \epsilon) = \epsilon R_n'(\xi), \tag{3.6.17}$$

with ξ on the line segment between $\widehat{\alpha} + \epsilon$ and $\widehat{\alpha} + 2\epsilon$, using the mean value theorem. Since

$$\begin{aligned}
\sup_{\xi \in [\widehat{\alpha}+\epsilon, \widehat{\alpha}+2\epsilon]} R_n'(\xi) &= \sup_{\xi \in [\widehat{\alpha}+\epsilon, \widehat{\alpha}+2\epsilon]} \left(\frac{\sum_{t=1}^n \overline{w}_t - n}{1-\xi} + \frac{\sum_{t=1}^n \overline{w}_t}{\xi} \right) \\
&= \sup_{\xi \in [\widehat{\alpha}+\epsilon, \widehat{\alpha}+2\epsilon]} \frac{-n\xi + \sum_{t=1}^n \overline{w}_t}{(1-\xi)\xi} \\
&= \sup_{\xi \in [\widehat{\alpha}+\epsilon, \widehat{\alpha}+2\epsilon]} \frac{n(\widehat{\alpha} - \xi)}{(1-\xi)\xi} \overset{n\to\infty}{\longrightarrow} -\infty \quad a.s.
\end{aligned}$$

and

$$p(\alpha) - p(\widehat{\alpha}) = p(\alpha) - p(\alpha_j) + o_P(1) = O_P(1) \tag{3.6.18}$$

which follows from $\widehat{\alpha} - \alpha_j = o_P(1)$ (due to the triangular inequality and (3.6.16)) and $p(\alpha)$ is a continuous function we have

$$H_n(\alpha) - H_n(\widehat{\alpha}) = R_n(\alpha) - R_n(\widehat{\alpha}) + p(\alpha) - p(\widehat{\alpha}) \overset{n\to\infty}{\longrightarrow} -\infty \tag{3.6.19}$$

in probability uniformly for $\alpha \in [\widehat{\alpha} + 2\epsilon, 1]$ which is a contradiction to $\overline{\alpha}^*$ being the maximizer of $H_n(\alpha)$. Therefore, we have $\overline{\alpha}^* < \widehat{\alpha} + 2\epsilon$ in probability. The analogous argumentation holds true to show that $\overline{\alpha}^* > \widehat{\alpha} + 2\epsilon$ in probability. Altogether, we have $\overline{\alpha}^* - \widehat{\alpha} = o_P(1)$, as claimed in (3.6.15).

By the definition $\overline{\alpha}^* = \arg\max_{\alpha \in [0,1]} H_n(\alpha)$ we have

$$H_n(\widehat{\alpha}) = R_n(\widehat{\alpha}) + p(\widehat{\alpha}) \leq H_n(\overline{\alpha}^*) = R_n(\overline{\alpha}^*) + p(\overline{\alpha}^*).$$

Applying a first order Taylor expansion at $\widehat{\alpha}$ for $R_n(\overline{\alpha}^*)$, we get

$$R_n(\overline{\alpha}^*) = R_n(\widehat{\alpha}) + R_n'(\widehat{\alpha})(\overline{\alpha}^* - \widehat{\alpha}) + \frac{R_n''(\eta)}{2}(\overline{\alpha}^* - \widehat{\alpha})^2$$

with η lying on the line segment between $\overline{\alpha}^*$ and $\widehat{\alpha}$. As noted at the beginning of the proof $\widehat{\alpha}$ is the maximizer of R_n implying $R_n'(\widehat{\alpha}) = 0$. Therefore, we get

$$R_n(\widehat{\alpha}) + p(\widehat{\alpha}) \leq R_n(\widehat{\alpha}) + \frac{R_n''(\eta)}{2}(\overline{\alpha}^* - \widehat{\alpha})^2 + p(\overline{\alpha}^*)$$

which is equivalent to

$$-\frac{R_n''(\eta)}{2}(\overline{\alpha}^* - \widehat{\alpha})^2 \leq p(\overline{\alpha}^*) - p(\widehat{\alpha}) = o_P(1). \tag{3.6.20}$$

Thereby, the last equality follows by (3.6.15) and the fact that $p(\alpha)$ is a continuous function in α. Note that

$$|R_n''(\eta)| = n\left\{\frac{1-\widehat{\alpha}}{(1-\eta)^2} + \frac{\widehat{\alpha}}{\eta^2}\right\}$$

By $\eta - \alpha_j = o_P(1)$ and $\widehat{\alpha} - \alpha_j = o_P(1)$ we get

$$|R_n''(\eta)| = \frac{n}{\alpha_j(1-\alpha_j)}\{1 + o_P(1)\} = O_P(n).$$

Inserting this into (3.6.20) we get

$$\overline{\alpha}^* - \widehat{\alpha} = o_P(n^{-1/2})$$

which implies $\overline{\alpha}^* - \widehat{\alpha} = O_P(n^{-1/4})$.

□

Lemma 3.5. *Let $(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma})$ be an EM-estimator of $(\alpha, \zeta_1, \zeta_2, \phi, \sigma)$. Under the null model the following holds:*

(i) If $\overline{\alpha} - 0.5 = O_P(n^{-1/4})$, then

$$\begin{aligned} &2\{pl_n(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}) - pl_n(0.5, 0, 0, \phi_0, 1)\} \\ \leq\ & \frac{(\sum_{t=1}^n Y_t)^2}{\sum_{t=1}^n Y_t^2} + \frac{(\sum_{t=1}^n Z_t)^2}{\sum_{t=1}^n Z_t^2} + \frac{\{(\sum_{t=1}^n V_t)^-\}^2}{\sum_{t=1}^n V_t^2} + \frac{(\sum_{t=1}^n W_{1t})^2}{\sum_{t=1}^n W_{1t}^2} + o_P(1), \end{aligned}$$

where x^- denotes the negative part of a real number x.

(ii) If $\overline{\alpha} - \alpha_j = o_P(1)$ for some $\alpha_j \in (0, 0.5)$, then

$$\begin{aligned} &2\{pl_n(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}) - pl_n(0.5, 0, 0, \phi_0, 1)\} \\ \leq\ & \frac{(\sum_{t=1}^n Y_t)^2}{\sum_{t=1}^n Y_t^2} + \frac{(\sum_{t=1}^n Z_t)^2}{\sum_{t=1}^n Z_t^2} + \frac{(\sum_{t=1}^n U_t)^2}{\sum_{t=1}^n U_t^2} + \frac{(\sum_{t=1}^n W_{1t})^2}{\sum_{t=1}^n W_{1t}^2} \\ & + 2\{p(\alpha_j) - p(0.5)\} + o_P(1). \end{aligned}$$

Proof. (*i*) Plugging (3.6.5) into the definition of $\overline{s}_3$ we get

$$\overline{s}_3 = \frac{\overline{\alpha}(1-2\overline{\alpha})}{(1-\overline{\alpha})^2}\overline{\zeta}_2^3 + \frac{\left(3\overline{\alpha}^2\overline{\zeta}_2^2 - 3\overline{s}_1\overline{\alpha}\overline{\zeta}_2 + \overline{s}_1^2\right)}{(1-\overline{\alpha})^2}\overline{s}_1,$$

which leads to

$$\overline{s}_3 = \frac{\overline{\alpha}(1-2\overline{\alpha})}{(1-\overline{\alpha})^2}\overline{\zeta}_2^3 + o_P(|\overline{s}_1|), \tag{3.6.21}$$

due to Lemma 3.1 and the assumption $\overline{\alpha} - 0.5 = o_P(1)$. For $\overline{\alpha} - 0.5 = O_P(n^{-1/4})$ we get

$$\overline{s}_3 = O_P(n^{-1/4})O_P(n^{-3/8}) + o_P(|\overline{s}_1|)$$

since $\overline{\zeta}_2 = O_P(n^{-1/8})$ (see Lemma 3.2). This entails $\overline{s}_3 = o_P(n^{-1/2})$ since $\overline{s}_1 = O_P(n^{-1/2})$ and therefore the third term in the expansion (3.6.8) is $o_P(1)$ and can be neglected, asymptotically.

In a next step we show that $\overline{s}_4$ is non-positive in probability. From equation (3.6.6) we know

$$\overline{s}_4 = \frac{\overline{\alpha}(1-6\overline{\alpha}+6\overline{\alpha}^2)}{(1-\overline{\alpha})^3}\overline{\zeta}_2^4 + o_P(|\overline{s}_1|). \tag{3.6.22}$$

By a zero addition we get

$$\begin{aligned}
\overline{s}_4 = & \frac{\overline{\alpha}(1-6\overline{\alpha}+6\overline{\alpha}^2)}{(1-\overline{\alpha})^3}\overline{\zeta}_2^4 - 3(1-2\overline{\alpha})\overline{\zeta}_2^2\overline{s}_3/(2(1-\overline{\alpha})) \\
& + 3(1-2\overline{\alpha})\overline{\zeta}_2^2\overline{s}_3/(2(1-\overline{\alpha})) + o_P(|\overline{s}_1|) \\
= & -\frac{\overline{\alpha}}{2(1-\overline{\alpha})^3}\overline{\zeta}_2^4 \\
& + \frac{3}{2}\overline{\zeta}_2\frac{6\overline{\zeta}_2^2\overline{\alpha}^3 - 6\overline{\zeta}_2\overline{s}_1\overline{\alpha}^2 - 3\overline{\zeta}_2^2\overline{\alpha}^2 + 3\overline{\zeta}_2\overline{s}_1\overline{\alpha} + 2\overline{\alpha}\overline{s}_1^2 - \overline{s}_1^2}{(1-\overline{\alpha})^3}\overline{s}_1 \\
& - \frac{3}{2}\frac{\overline{\zeta}_2(2\overline{\alpha}-1)}{1-\overline{\alpha}}\overline{s}_3 + o_P(|\overline{s}_1|)
\end{aligned} \tag{3.6.23}$$

which leads to (using Lemma 3.1)

$$\overline{s}_4 = -\frac{\overline{\alpha}}{2(1-\overline{\alpha})^3}\overline{\zeta}_2^4 + o_P(|\overline{s}_1|) + o_P(|\overline{s}_3|).$$

Due to (3.6.9) we have $\overline{s}_1 = O_P(n^{-1/2})$ and $\overline{s}_3 = O_P(n^{-1/2})$ and thus

$$\overline{s}_4 = -\frac{\overline{\alpha}}{2(1-\overline{\alpha})^3}\overline{\zeta}_2^4 + o_P(n^{-1/2}).$$

Note that $\overline{\alpha} - 0.5 = O_P(n^{-1/4})$ by assumption and therefore $-\frac{\overline{\alpha}}{2(1-\overline{\alpha})^3} = -2 + o_P(1)$ leading to

$$\overline{s}_4 = -2\overline{\zeta}_2^4 + o_P(n^{-1/2}) \tag{3.6.24}$$

since $\overline{\zeta}_2^4 = O_P(n^{-1/2})$. Hence, we can strengthen the upper bound in (3.6.8) to

$$\begin{aligned} & 2\{pl_n(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}) - pl_n(0.5, 0, 0, \phi_0, 1)\} \\ \leq\ & 2\{\overline{s}_1 \sum_{t=1}^n Y_t + \overline{s}_2 \sum_{t=1}^n Z_t + \overline{s}_4 \sum_{t=1}^n V_t + \overline{s}_5 \sum_{t=1}^n W_{1t}\} \qquad (3.6.25) \\ & -\{\overline{s}_1^2 \sum_{t=1}^n Y_t^2 + \overline{s}_2^2 \sum_{t=1}^n Z_t^2 + \overline{s}_4^2 \sum_{t=1}^n V_t^2 + \overline{s}_5^2 \sum_{t=1}^n W_{1t}^2\}\{1 + o_p(1)\} + o_P(1) \\ \leq\ & \frac{(\sum_{t=1}^n Y_t)^2}{\sum_{t=1}^n Y_t^2} + \frac{(\sum_{t=1}^n Z_t)^2}{\sum_{t=1}^n Z_t^2} + \frac{\{(\sum_{t=1}^n V_t)^-\}^2}{\sum_{t=1}^n V_t^2} + \frac{(\sum_{t=1}^n W_{1t})^2}{\sum_{t=1}^n W_{1t}^2} + o_P(1). \qquad (3.6.26) \end{aligned}$$

For the last inequality we used the property of quadratic functions, $\overline{s}_4 \leq 0$ in probability and $\overline{\alpha} - 0.5 = o_P(1)$ together with $p(\alpha)$ being a continuous function in α.

(ii) Plugging (3.6.5) into the definition of $\overline{s}_3$ we get

$$\overline{s}_3 = \frac{\overline{\alpha}(1 - 2\overline{\alpha})}{(1 - \overline{\alpha})^2} \overline{\zeta}_2^3 + o_P(|\overline{s}_1|)$$

as in the proof of (i). Due to the order information in Lemma 3.2, we get $\overline{\zeta}_2 = O_P(n^{-1/6})$. By symmetry, $\overline{\zeta}_1 = O_P(n^{-1/6})$. From the definition of $\overline{s}_2$ we conclude $\overline{\sigma}^2 - 1 = O_P(n^{-1/3})$. From (3.6.24) we get $\overline{s}_4 = o_P(n^{-1/2})$. Therefore, the fourth term in the expansion (3.6.8) is $o_P(1)$ and can be neglected asymptotically and we strengthen the upper bound in (3.6.8) to

$$\begin{aligned} & 2\{pl_n(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}) - pl_n(0.5, 0, 0, \phi_0, 1)\} \\ \leq\ & 2\{\overline{s}_1 \sum_{t=1}^n Y_t + \overline{s}_2 \sum_{t=1}^n Z_t + \overline{s}_3 \sum_{t=1}^n U_t + \overline{s}_5 \sum_{t=1}^n W_{1t}\} \\ & -\{\overline{s}_1^2 \sum_{t=1}^n Y_t^2 + \overline{s}_2^2 \sum_{t=1}^n Z_t^2 + \overline{s}_3^2 \sum_{t=1}^n U_t^2 + \overline{s}_5^2 \sum_{t=1}^n W_{1t}^2\}\{1 + o_p(1)\} \qquad (3.6.27) \\ & +2\{p(\alpha_j) - p(0.5)\} + o_P(1) \\ \leq\ & \frac{(\sum_{t=1}^n Y_t)^2}{\sum_{t=1}^n Y_t^2} + \frac{(\sum_{t=1}^n Z_t)^2}{\sum_{t=1}^n Z_t^2} + \frac{(\sum_{t=1}^n U_t)^2}{\sum_{t=1}^n U_t^2} + \frac{(\sum_{t=1}^n W_{1t})^2}{\sum_{t=1}^n W_{1t}^2} \qquad (3.6.28) \\ & +2\{p(\alpha_j) - p(0.5)\} + o_P(1). \end{aligned}$$

□

Remark 3.3. If the order of the AR(p) process under the null model is higher than 1; then the upper bounds in Lemma 3.5 for

$$2\{pl_n(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}) - pl_n(0.5, 0, 0, \phi_0, 1)\}$$

will be the same up to replacing

$$\frac{(\sum_{t=1}^n W_{1t})^2}{\sum_{t=1}^n W_{1t}^2} \text{ by } \sum_{\tau=1}^{p}\left(\frac{(\sum_{t=1}^n \tilde{W}_{\tau t})^2}{\sum_{t=1}^n \tilde{W}_{\tau t}^2}\right). \tag{3.6.29}$$

Note that these terms will also occur in (3.6.30) and cancel out when considering

$$2\{pl_n(\overline{\alpha},\overline{\zeta}_1,\overline{\zeta}_2,\overline{\phi},\overline{\sigma}) - pl_n(0.5,\widehat{\zeta}_0,\widehat{\zeta}_0,\widehat{\phi}_0,\widehat{\sigma}_0)\},$$

though.

Proof of Theorem 3.1. We know that

$$\begin{aligned}
&2\{pl_n(0.5,\widehat{\zeta}_0,\widehat{\zeta}_0,\widehat{\phi}_0,\widehat{\sigma}_0) - pl_n(0.5,0,0,\phi_0,1)\}\\
=&2\{l_n(0.5,\widehat{\zeta}_0,\widehat{\zeta}_0,\widehat{\phi}_0,\widehat{\sigma}_0) - l_n(0.5,0,0,\phi_0,1)\}\\
=&\frac{(\sum_{t=1}^n Y_t)^2}{\sum_{t=1}^n Y_t^2} + \frac{(\sum_{t=1}^n Z_t)^2}{\sum_{t=1}^n Z_t^2} + \frac{(\sum_{t=1}^n W_{1t})^2}{\sum_{t=1}^n W_{1t}^2} + o_P(1).
\end{aligned} \tag{3.6.30}$$

Using the results of Lemma 3.5 we get

$$M_n^{(K)}(0.5) \le \frac{\{(\sum_{t=1}^n V_t)^-\}^2}{\sum_{t=1}^n V_t^2} + o_P(1)$$

and

$$M_n^{(K)}(\alpha_j) \le \frac{(\sum_{t=1}^n U_t)^2}{\sum_{t=1}^n U_t^2} + 2\{p(\alpha_j) - p(0.5)\} + o_P(1)$$

for $\alpha_j \ne 0.5$. Note that this inequality still holds true if we replace $2\{p(\alpha_j) - p(0.5)\}$ by $\Delta = 2\max_{\alpha_j\ne 0.5}\{p(\alpha_j) - p(0.5)\}$ as defined in Theorem 3.1. Therefore,

$$EM_n^{(K)} \le \max\left[\frac{(\sum_{t=1}^n U_t)^2}{\sum_{t=1}^n U_t^2} + \Delta, \frac{\{(\sum_{t=1}^n V_t)^-\}^2}{\sum_{t=1}^n V_t^2}\right] + o_P(1).$$

Our next step is to show that this upper bound is even attained. Note that due to the *EM-property* we have

$$M_n^{(K)}(\alpha_j) \ge M_n^{(0)}(\alpha_j), \quad j = 1,\ldots,J. \tag{3.6.31}$$

We distinguish the two cases $\alpha_j = 0.5$ and $\alpha_j \ne 0.5$.

First, let $\alpha_j = 0.5$. Since we have (3.6.31) it suffices to find values $(\zeta_1,\zeta_2,\phi,\sigma)$ such that for fixed $\alpha_j = 0.5$ the upper bound in Lemma 3.5 is attained. Therefore, we have to find $s_j = \hat{s}_j + o_P(n^{-1/2})$, $j = 1,2,4,5$, where

$$\hat{s}_1 = \frac{\sum_{t=1}^n Y_t}{\sum_{t=1}^n Y_t^2},\ \hat{s}_2 = \frac{\sum_{t=1}^n Z_t}{\sum_{t=1}^n Z_t^2},\ \hat{s}_4 = \frac{(\sum_{t=1}^n V_t)^-}{\sum_{t=1}^n V_t^2},\ \hat{s}_5 = \frac{\sum_{t=1}^n W_{1t}}{\sum_{t=1}^n W_{1t}^2},$$

and s_j's defined as in the proof of Lemma 3.2. Neglecting terms of order $o_P(n^{-1/2})$ we are searching for ζ_1, ζ_2, σ and ϕ satisfying

$$\begin{aligned} \hat{s}_1 &= \frac{1}{2}(\zeta_1 + \zeta_2), \\ \hat{s}_2 &= \frac{1}{2}(\zeta_1^2 + \zeta_2^2) + (\sigma^2 - 1), \\ \hat{s}_4 &= -2\zeta_2^4, \\ \hat{s}_5 &= (\phi - \phi_0). \end{aligned} \tag{3.6.32}$$

Note here, that the equation $\hat{s}_4 = -2\zeta_2^4 + o_P(n^{-1/2})$ is due to (3.6.24). Without loss of generality, let $\tilde{\zeta}_2$ be the non-negative solution of $\hat{s}_4 = -2\zeta_2^4$. Using the conditions in (3.6.32) we get

$$\begin{aligned} \tilde{\zeta}_1 &= 2\hat{s}_1 - \tilde{\zeta}_2, \\ \tilde{\sigma}^2 &= \hat{s}_2 - \frac{1}{2}(\tilde{\zeta}_1^2 + \tilde{\zeta}_2^2) + 1 \end{aligned}$$

and

$$\tilde{\phi} = \hat{s}_5 + \phi_0.$$

One immediately sees that $\tilde{s}_j = \hat{s}_j + o_P(n^{-1/2})$, $j = 1, 2, 4, 5$. Here, $\hat{s}_5 = O_P(n^{-1/2})$ follows by the CLT for stationary and ergodic martingale differences (applied to $\sum_{t=1}^n W_{1t}$) and ergodic theorem (applied to $\sum_{t=1}^n W_{1t}^2$). Using $\hat{s}_j = O_P(n^{-1/2})$, $j = 4, 5$, we get $\tilde{\phi} - \phi_0 = O_P(n^{-1/2})$ and $\tilde{\zeta}_2 = O_P(n^{-1/8})$. By symmetry, $\tilde{\zeta}_1 = O_P(n^{-1/8})$. Since $\tilde{\zeta}_1 = O_P(n^{-1/8})$, $\tilde{\zeta}_2 = O_P(n^{-1/8})$ and $\hat{s}_2 = O_P(n^{-1/2})$ we have $\tilde{\sigma}^2 - 1 = O_P(n^{-1/4})$.

Putting these values $\tilde{s}_j$, $j = 1, 2, 3, 4$ into (3.6.25), we see that the upper bound (3.6.26) is also attained, i.e.

$$\begin{aligned} & 2\{pl_n(0.5, \tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\phi}, \tilde{\sigma}) - pl_n(0.5, 0, 0, \phi_0, 1)\} \\ = \; & \frac{(\sum_{t=1}^n Y_t)^2}{\sum_{t=1}^n Y_t^2} + \frac{(\sum_{t=1}^n Z_t)^2}{\sum_{t=1}^n Z_t^2} + \frac{\{(\sum_{t=1}^n V_t)^-\}^2}{\sum_{t=1}^n V_t^2} + \frac{(\sum_{t=1}^n W_{1t})^2}{\sum_{t=1}^n W_{1t}^2} + o_P(1), \end{aligned}$$

and therefore

$$2\{pl_n(0.5, \tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\phi}, \tilde{\sigma}) - pl_n(0.5, \widehat{\zeta}_0, \widehat{\zeta}_0, \widehat{\phi}_0, \widehat{\sigma}_0)\} = \frac{\{(\sum_{t=1}^n V_t)^-\}^2}{\sum_{t=1}^n V_t^2} + o_P(1).$$

Since we have (3.6.31), i.e. we have the so called *EM-property*, we know that

$$M_n^{(K)}(0.5) \geq M_n^{(0)}(0.5) = \frac{\{(\sum_{t=1}^n V_t)^-\}^2}{\sum_{t=1}^n V_t^2} + o_P(1).$$

This is why the upper bound is attained.

Second, let $\alpha_j \neq 0.5$. Since we have (3.6.31) it suffices to find values $(\zeta_1, \zeta_2, \phi, \sigma)$ so that for fixed $\alpha_j \neq 0.5$ the upper bound in Lemma 3.5 is attained. Let $\alpha_{j_0} \neq 0.5$ such that $\alpha_{j_0} = \arg\max_{\alpha_j \in \mathcal{J}\setminus\{0.5\}} 2\{p(\alpha_j) - p(0.5)\}$. Therefore, we have to find $s_j = \hat{s}_j + o_P(n^{-1/2})$, $j = 1, 2, 4, 5$ where

$$\hat{s}_1 = \frac{\sum_{t=1}^n Y_t}{\sum_{t=1}^n Y_t^2},\ \hat{s}_2 = \frac{\sum_{t=1}^n Z_t}{\sum_{t=1}^n Z_t^2},\ \hat{s}_3 = \frac{\sum_{t=1}^n U_t}{\sum_{t=1}^n U_t^2},\ \hat{s}_5 = \frac{\sum_{t=1}^n W_{1t}}{\sum_{t=1}^n W_{1t}^2},$$

and s_j's as in the proof of Lemma 3.2. Neglecting terms of order $o_P(n^{-1/2})$ we are searching for ζ_1, ζ_2, σ and ϕ satisfying

$$\begin{aligned} \hat{s}_1 &= (1-\alpha_{j_0})\zeta_1 + \alpha_{j_0}\zeta_2, \\ \hat{s}_2 &= (1-\alpha_{j_0})\zeta_1^2 + \alpha_{j_0}\zeta_2^2 + (\sigma^2 - 1), \\ \hat{s}_3 &= \frac{\alpha_{j_0}(1-2\alpha_{j_0})}{(1-\alpha_{j_0})^2}\zeta_2^3, \\ \hat{s}_5 &= (\phi - \phi_0). \end{aligned} \tag{3.6.33}$$

Note here that $\hat{s}_3 = \frac{\alpha_{j_0}(1-2\alpha_{j_0})}{(1-\alpha_{j_0})^2}\zeta_2^3 + o_P(n^{-1/2})$ is due to (3.6.21) and $\overline{s}_1 = O_P(n^{-1/2})$. Let $\tilde{\zeta}_2$ be the real solution of $\hat{s}_3 = \frac{\alpha_{j_0}(1-2\alpha_{j_0})}{(1-\alpha_{j_0})^2}\zeta_2^3$. Using the conditions in (3.6.33),

$$\begin{aligned} \tilde{\zeta}_1 &= \frac{\hat{s}_1 - \alpha_{j_0}\tilde{\zeta}_2}{1-\alpha_{j_0}}, \\ \tilde{\sigma}^2 &= \hat{s}_2 - ((1-\alpha_{j_0})\tilde{\zeta}_1^2 + \alpha_{j_0}\tilde{\zeta}_2^2) + 1, \end{aligned}$$

and

$$\tilde{\phi} = \hat{s}_5 + \phi_0.$$

One immediately sees that $\tilde{s}_j = \hat{s}_j + o_P(n^{-1/2})$, $j = 1, 2, 4, 5$. Using $\hat{s}_j = O_P(n^{-1/2})$ it is $\tilde{\phi} - \phi_0 = O_P(n^{-1/2})$ and $\tilde{\zeta}_2 = O_P(n^{-1/6})$. By the first equation in (3.6.32) and $\tilde{\zeta}_2 = O_P(n^{-1/6})$ we get $\tilde{\zeta}_1 = O_P(n^{-1/6})$. Finally, $\tilde{\sigma}^2 - 1 = O_P(n^{-1/3})$ by $\tilde{\zeta}_i = O_P(n^{-1/6})$, $i = 1, 2$, and the second equation in (3.6.33).

Putting these values $\tilde{s}_j$, $j = 1, 2, 3, 5$ into (3.6.27), we see that the upper bound (3.6.28) is also attained, i.e.

$$\begin{aligned} & 2\{pl_n(\alpha_{j_0}, \tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\phi}, \tilde{\sigma}) - pl_n(0.5, 0, 0, \phi_0, 1)\} \\ = & \frac{(\sum_{t=1}^n Y_t)^2}{\sum_{t=1}^n Y_t^2} + \frac{(\sum_{t=1}^n Z_t)^2}{\sum_{t=1}^n Z_t^2} + \frac{(\sum_{t=1}^n U_t)^2}{\sum_{t=1}^n U_t^2} + \frac{(\sum_{t=1}^n W_{1t})^2}{\sum_{t=1}^n W_{1t}^2} \\ & + \Delta + o_P(1), \end{aligned}$$

and thus

$$2\{pl_n(\alpha_{j_0},\tilde{\zeta}_1,\tilde{\zeta}_2,\tilde{\phi},\tilde{\sigma})-pl_n(0.5,\widehat{\zeta}_0,\widehat{\zeta}_0,\widehat{\phi}_0,\widehat{\sigma}_0)\}=\frac{(\sum_{t=1}^n U_t)^2}{\sum_{t=1}^n U_t^2}+\Delta+o_P(1).$$

By the *EM-property* we know that

$$M_n^{(K)}(\alpha_{j_0})\geq M_n^{(0)}(\alpha_{j_0})=\frac{(\sum_{t=1}^n U_t)^2}{\sum_{t=1}^n U_t^2}+\Delta+o_P(1),\quad j_0\in\{1,\ldots,J\}.$$

Hence, the upper bound is attained. Altogether, we have

$$EM_n^{(K)}=\max\left[\frac{(\sum_{t=1}^n U_t)^2}{\sum_{t=1}^n U_t^2}+\Delta,\frac{\{(\sum_{t=1}^n V_t)^-\}^2}{\sum_{t=1}^n V_t^2}\right]+o_P(1).$$

By the multivariate central limit theorem $(1/\sqrt{n})\sum_{t=1}^n(U_t,V_t)^T$ is bivariate normal, asymptotically. Since U_t and V_t are uncorrelated $(1/\sqrt{n})\sum_{t=1}^n U_t$ and $(1/\sqrt{n})\sum_{t=1}^n V_t$ are asymptotically independent. Therefore, the limiting distribution is given by $F(x-\Delta)\{\mathbb{1}_{\{x\geq 0\}}+F(x)\}/2$, $x\in\mathbb{R}$, where $F(\cdot)$ is the cdf of a χ_1^2 variate. □

Remark 3.4. The only problem that occurs whenever the order of the AR(p) is higher than 1 is to show that the upper bound in Lemma 3.5 is attained. For this, we can proceed as in the case $p=1$ and are keen on finding $(\zeta_1,\zeta_2,\boldsymbol{\phi},\sigma)$, such that for fixed α_j this upper bound is attained. We choose $\tilde{\zeta}_1$, $\tilde{\zeta}_2$ and $\tilde{\sigma}$ as before. Choosing $\tilde{\boldsymbol{\phi}}$ is more complicated, though. To this end, let $\hat{s}'_{4+\tau}=\frac{\sum_{t=1}^n\tilde{W}_{\tau t}}{\sum_{t=1}^n\tilde{W}_{\tau t}^2}$, $\tau=1,\ldots,p$. Neglecting terms of order $o_P(n^{-1/2})$, we are searching for parameters $\phi_1,\ldots,\phi_p$ satisfying

$$\hat{s}'_{4+p}=s_{4+p},\quad \hat{s}'_{3+p}=\alpha_{1,p-1}s_{4+p}+\alpha_{2,p-1}s_{3+p},\quad \ldots,\quad \hat{s}'_5=\sum_{\tau=1}^p\alpha_{\tau,1}s_{4+\tau} \tag{3.6.34}$$

with $s_{4+\tau}=\phi_\tau-\phi_{\tau,0}$, $\tau=1,\ldots,p$, and coefficients $\alpha_{\tau,l}$, $l=1,\ldots,p-1$ and $\tau=1,\ldots,p-l+1$, determined by the orthogonalisation in (3.6.11) in Remark 3.2. Therefore, we choose

$$\tilde{\phi}_p=\hat{s}'_{4+p}+\phi_{p,0}.$$

Note that $\hat{s}'_{4+p}=O_P(n^{-1/2})$ (due to the ergodic theorem, applied to the denominator, and CLT for stationary and ergodic martingale differences, applied to the numerator) and so $\tilde{\phi}_p-\phi_{p,0}=O_P(n^{-1/2})$. Replacing s_{4+p} by $\tilde{\phi}_p-\phi_{p,0}$ in the second equation of (3.6.34), we get $\tilde{\phi}_{p-1}$. Further, since we have $\tilde{\phi}_p-\phi_{p,0}=O_P(n^{-1/2})$ and $\hat{s}'_{3+p}=O_P(n^{-1/2})$ (by the same reasoning as for $\hat{s}'_{4+p}$), we also have $\tilde{\phi}_{p-1}-\phi_{p-1,0}=O_P(n^{-1/2})$. Repeating this procedure we obtain $\tilde{\phi}_{p-2},\tilde{\phi}_{p-3},\ldots,\tilde{\phi}_1$.

3.6.1 Orthogonality of Y_t, Z_t, U_t, V_t and $W_{\tau t}$.

Let $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots)$ be the filtration generated by the observations $(X_k)_k$. For $k \in \mathbb{N}$ we have

$$E\epsilon_1^{2k-1} = 0 \text{ and } E\epsilon_1^{2k} = 1 \cdot 3 \cdot \ldots \cdot (2k-1)$$

and therefore

$$\begin{aligned}
E[Y_1 Z_1] &= \frac{1}{2} E[\epsilon_1(\epsilon_1^2 - 1)] \\
&= \frac{1}{2} E[\epsilon_1^3 - \epsilon_1] = 0, \\
E[Y_1 U_1] &= \frac{1}{6} E[\epsilon_1(\epsilon_1^3 - 3\epsilon_1)] \\
&= \frac{1}{6} E[\epsilon_1^4 - 3\epsilon_1^2] = \frac{1}{6}(3-3) = 0, \\
E[Y_1 V_1] &= \frac{1}{24} E[\epsilon_1(\epsilon_1^4 - 6\epsilon_1^2 + 3)] \\
&= \frac{1}{24} E[\epsilon_1^5 - 6\epsilon_1^3 + 3\epsilon_1] = 0.
\end{aligned}$$

For $\tau = 1, \ldots, p$,

$$\begin{aligned}
E[Y_1 W_{\tau 1}] &= E[X_{1-\tau} Y_1^2] \\
&= E[X_{1-\tau}\epsilon_1^2] \\
&= E[E[X_{1-\tau}\epsilon_1^2 | \mathcal{F}_0]] \\
&= E[X_{1-\tau} \underbrace{E[\epsilon_1^2 | \mathcal{F}_0]}_{=1}] \\
&= E[X_{1-\tau}] \\
&= 0,
\end{aligned}$$

since we assumed $(X_k)_k$ to be a causal AR(p) process under the null with intercept $\zeta_0 = 0$. Further,

$$\begin{aligned}
E[Z_1 U_1] &= \frac{1}{12} E[(\epsilon_1^2 - 1)(\epsilon_1^3 - 3\epsilon_1)] \\
&= \frac{1}{12} E[\epsilon_1^5 - 4\epsilon_1^3 + 3\epsilon_1] = 0, \\
E[Z_1 V_1] &= \frac{1}{48} E[(\epsilon_1^2 - 1)(\epsilon_1^4 - 6\epsilon_1^2 + 3)] \\
&= \frac{1}{48} E[\epsilon_1^6 - 7\epsilon_1^4 + 9\epsilon_1^2 - 3] = \frac{1}{48}(15 - 21 + 9 - 3) = 0.
\end{aligned}$$

For $\tau = 1, \dots, p$,

$$E[Z_1 W_{\tau 1}] = E[X_{1-\tau} Y_1 Z_1] = E[X_{1-\tau} E[Y_1 Z_1 | \mathcal{F}_0]] = 0$$

since $E[Y_1 Z_1 | \mathcal{F}_0] = 0$.
Since $E\epsilon_1^{2k-1} = 0$ for $k \in \mathbb{N}$,

$$\begin{aligned} E[U_1 V_1] &= \frac{1}{144} E[(\epsilon_1^3 - 3\epsilon_1)(\epsilon_1^4 - 6\epsilon_1^2 + 3)] \\ &= \frac{1}{144} E[\epsilon_1^7 - 9\epsilon_1^5 + 21\epsilon_1^3 - 9\epsilon_1] = 0. \end{aligned}$$

For $\tau = 1, \dots, p$,

$$\begin{aligned} E[U_1 W_{\tau 1}] &= E[X_{1-\tau} U_1 Y_1] \\ &= E[X_{1-\tau} E[U_1 Y_1 | \mathcal{F}_0]] = 0 \end{aligned}$$

since $E[U_1 Y_1 | \mathcal{F}_0] = 0$.
For $\tau = 1, \dots, p$,

$$\begin{aligned} E[V_1 W_{\tau 1}] &= E[X_{1-\tau} V_1 Y_1] \\ &= E[X_{1-\tau} E[V_1 Y_1 | \mathcal{F}_0]] \\ &= 0 \end{aligned}$$

since $E[V_1 Y_1 | \mathcal{F}_0] = 0$.
Therefore, we know that $W_{\tau 1}$, $\tau = 1, \dots, p$ is orthogonal to Y_1, Z_1, U_1 and V_1 but not to $W_{\tau' 1}$, $\tau' \neq \tau$ as can be seen by

$$\begin{aligned} E[W_{\tau 1} W_{\tau' 1}] &= E[X_{1-\tau} X_{1-\tau'} Y_1^2] \\ &= E[X_{1-\tau} X_{1-\tau'} \epsilon_1^2] \\ &= E[E[X_{1-\tau} X_{1-\tau'} \epsilon_1^2 | \mathcal{F}_0]] \\ &= E[X_{1-\tau} X_{1-\tau'} \underbrace{E[\epsilon_1^2 | \mathcal{F}_0]}_{=1}] \\ &= E[X_{1-\tau} X_{1-\tau'}] \\ &= \gamma(|\tau - \tau'|) \neq 0, \end{aligned}$$

where $\gamma(\cdot)$ is the autocovariance function of the (stationary) process $(X_k)_k$ since we assume $\zeta_0 = 0$ implying $EX_0 = 0$.

4 Testing in a Markov-switching intercept-variance model

In this chapter we discuss testing for homogeneity in a linear switching autoregressive model with switching intercept and switching scale parameter of the normal innovations. For mixture models Chen and Li (2009) recently developed the so called EM-test for testing for homogeneity in a normal mixture model with possibly distinct means and variances under the alternative. When compared to testing for homogeneity in homoscedastic normal mixture models, the asymptotic properties of likelihood based methods become much more challenging which is due to unbounded (log) likelihood and possibly infinite Fisher information, see Chen and Li (2009). The problem of unbounded (log) likelihood arises if one sets one of the location parameters, say μ_1, equal to an observation and the associated standard deviation σ_1 tends to 0. Several formal ways around this problem have been suggested in the literature, e.g. choose the largest local maximum or restrict the possible variances by restrictions of the form $\sigma_1^2 \leq c\sigma_2^2$ and $\sigma_2^2 \leq c\sigma_1^2$, for some $c > 1$, cf. Hathaway (1985). Chen and Li (2009) introduce a penalty function on σ such that the estimators for σ_1 and σ_2 are forced away from zero. They show that their EM-test admits a rather simple asymptotic distribution. In this chapter we extend this EM-test to switching autoregressive models with switching intercept and switching variance under the alternative and normally distributed innovations .

4.1 Testing in a linear switching autoregressive model with possibly switching intercept and variance

The linear switching autoregressive model with possibly switching intercept and scale parameter under the alternative is given by

$$X_t = \zeta_{S_t} + \sum_{j=1}^{p} \phi_j X_{t-j} + \sigma_{S_t}\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0,1), \tag{4.1.1}$$

where the ϕ_j's are the (non-switching) autoregressive parameters, and the intercept ζ as well as the scale parameter σ switch according to S_t.

Writing model (4.1.1) for a two-state chain $(S_k)_k$ of the form $X_t = F_{\boldsymbol{\omega}}(S_t, X_{t-1}^p; \epsilon_t)$ with finite dimensional parameter $\boldsymbol{\omega} = (a_{12}, a_{21}, \vartheta_1, \vartheta_2, \boldsymbol{\eta}^T)^T$, we have $\vartheta_i = (\zeta_i, \sigma_i)^T$, $i = 1, 2$, and $\boldsymbol{\eta} = (\phi_1, \ldots, \phi_p)^T$. This is in sharp contrast to the previous chapters where we allowed for a univariate switching parameter. Throughout this chapter we assume $(\zeta, \sigma) \in Z \times [\delta, \infty) = \Theta \subset \mathbb{R}^2$, $\delta > 0$, and $\boldsymbol{\eta} \in \boldsymbol{H}$ where $\boldsymbol{H}$ and Z are any subsets of $\mathbb{R}^p$ and $\mathbb{R}$.

A special case of model (4.1.1) has been considered in Velucchi (2009). She fitted an HMM with two states and sdfs $P(X_t \leq x | S_t = i) = \Phi\big((x - \zeta_i)/\sigma_i\big)$, $i = 1, 2$, to Italian stock market returns. She showed that there exist two regimes: one regime with low returns and high volatility and a second regime with high returns and low volatility.

A slightly different version of model (4.1.1) can be found in Bhar and Hamori (2004). They developed their so called Markov-Switching Stock Return Model for modeling stock returns

$$X_t - \mu_{S_t} = \phi(X_{t-1} - \mu_{S_{t-1}}) + \sigma_{S_t}\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0, 1). \tag{4.1.2}$$

In model (4.1.1) a level shift in the mean occurs immediately when changing the state of the hidden Markov chain while the mean level in model (4.1.2) approaches the new value smoothly over several periods. Note here, that model (4.1.1) with $p = 1$ and (4.1.2) are equivalent if the hidden Markov chain consists of just one state.

The aim of this chapter is to give a feasible method for testing the null hypothesis $\mathcal{M} = \{1\}$ of a single regime against the alternative $\mathcal{M} = \{1, 2\}$ of (at least) two regimes.
To ensure identifiability of the parameters and for the uniqueness of the order p, we suppose that under the null model, i.e. no regime switch, $(X_k)_k$ is a causal AR(p) process.

Quasi-Likelihood-Ratio

As in Cho and White (2007) we consider a model which ignores the serial correlation in $(S_k)_k$ but captures the serial correlation of the process $(X_k)_k$. Even if we ignore the serial correlation in $(S_k)_k$ we are able to test for the number of regimes. Note here, that we allow two parameters to switch while Cho and White (2007) confine their considerations to a univariate switching parameter.

Let $X_1, \ldots, X_n$ be a random sample of size n from model (4.1.1). We do not work with the (full) likelihood conditional on the intial observations $(X_0, \ldots, X_{-p+1})$ and the initial state $S_0 = i_0$. Instead, we consider the quasi log-likelihood function which is given by

$$l_n(\boldsymbol{\psi}) = \sum_{t=1}^{n} \log\big((1-\alpha) g(X_t | X_{t-1}^p; \zeta_1, \boldsymbol{\phi}, \sigma_1) + \alpha g(X_t | X_{t-1}^p; \zeta_2, \boldsymbol{\phi}, \sigma_2)\big). \tag{4.1.3}$$

where $\boldsymbol{\psi} = (\alpha, \zeta_1, \zeta_2, \boldsymbol{\phi}^T, \sigma_1, \sigma_2)^T$ and $(1-\alpha, \alpha)$ corresponds to the stationary distribution of the hidden Markov chain $(S_k)_k$. Assuming that the innovations $(\epsilon_k)_k$ are independent and identically normally distributed, the conditional density (w.r.t. Lebesgue measure on $\mathbb{R}$) of X_t given $X_{t-1}^p = x_{t-1}^p$ and $S_t = i$ is given by

$$g(x_t|x_{t-1}^p;\zeta_i,\boldsymbol{\phi},\sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(-\frac{x_t-\zeta_i-\sum_{j=1}^p\phi_j x_{t-j}}{2\sigma_i^2}\right).$$

Our aim is to test the hypothesis of no regime switch, i.e.

$$H:\alpha(1-\alpha)=0 \text{ or } (\zeta_1,\sigma_1)=(\zeta_2,\sigma_2).$$

4.2 The EM-test

Similar to Chen et al. (2001, 2004) we consider a modified quasi log likelihood function which is defined by:

$$pl_n(\alpha,\zeta_1,\zeta_2,\boldsymbol{\phi},\sigma_1,\sigma_2) = l_n(\alpha,\zeta_1,\zeta_2,\boldsymbol{\phi},\sigma_1,\sigma_2)+p(\alpha)$$

where $p(\alpha)$ is a penalty function on α and fulfills the properties given in Section 1.5.

In the following we describe the (quasi) EM-test for testing for homogeneity in model (4.1.1). Note that in the following algorithm we proceed in some steps via the ECM algorithm instead of the EM algorithm. If we use the EM algorithm, we have to derive the updated estimators $(\zeta_{1j}^{(k+1)},\zeta_{2j}^{(k+1)},\boldsymbol{\phi}_j^{(k+1)},\sigma_{1j}^{(k+1)}\sigma_{2j}^{(k+1)})$ in Step 3 by maximizing

$$\sum_{t=1}^n(1-w_{tj}^{(k)})\log g(X_t|X_{t-1}^p;\zeta_1,\boldsymbol{\phi},\sigma_1)+\sum_{t=1}^n w_{tj}^{(k)}\log g(X_t|X_{t-1}^p;\zeta_2,\boldsymbol{\phi},\sigma_2)$$

simultaneously over $Z^2\times\boldsymbol{H}\times[\delta,\infty)^2$.

Step 0. Choose $0<\alpha_1<\alpha_2<\ldots<\alpha_J=0.5$. Compute

$$(\widehat{\zeta}_0,\widehat{\boldsymbol{\phi}}_0,\widehat{\sigma}_0)=\arg\max_{\zeta,\boldsymbol{\phi},\sigma} pl_n(0.5,\zeta,\zeta,\boldsymbol{\phi},\sigma,\sigma).$$

Put $j=1$ and $k=0$.

Step 1. Put $\alpha_j^{(k)}=\alpha_j$.

Step 2. Compute

$$(\zeta_{1j}^{(k)},\zeta_{2j}^{(k)},\boldsymbol{\phi}_j^{(k)},\sigma_{1j}^{(k)},\sigma_{2j}^{(k)})=\arg\max_{\zeta_1,\zeta_2,\boldsymbol{\phi},\sigma_1,\sigma_2} pl_n(\alpha_j^{(k)},\zeta_1,\zeta_2,\boldsymbol{\phi},\sigma_1,\sigma_2)$$

and

$$M_n^{(k)}(\alpha_j) = 2\left\{pl_n(\alpha_j^{(k)}, \zeta_{1j}^{(k)}, \zeta_{2j}^{(k)}, \boldsymbol{\phi}_j^{(k)}, \sigma_{1j}^{(k)}, \sigma_{2j}^{(k)}) - pl_n(0.5, \widehat{\zeta}_0, \widehat{\zeta}_0, \widehat{\boldsymbol{\phi}}_0, \widehat{\sigma}_0, \widehat{\sigma}_0)\right\}.$$

Step 3. Compute for $t = 1, \ldots, n$ the weights

$$w_{tj}^{(k)} = \frac{\alpha_j^{(k)} g(X_t|X_{t-1}^p; \zeta_{2j}^{(k)}, \boldsymbol{\phi}_j^{(k)}, \sigma_{2j}^{(k)})}{(1-\alpha_j^{(k)}) g(X_t|X_{t-1}^p; \zeta_{1j}^{(k)}, \boldsymbol{\phi}_j^{(k)}, \sigma_{1j}^{(k)}) + \alpha_j^{(k)} g(X_t|X_{t-1}^p; \zeta_{2j}^{(k)}, \boldsymbol{\phi}_j^{(k)}, \sigma_{2j}^{(k)})}.$$

Compute the estimators

$$\begin{aligned}
\alpha_j^{(k+1)} &= \arg\max_{\alpha} \left((n - \sum_{t=1}^n w_{tj}^{(k)}) \log(1-\alpha) + \sum_{t=1}^n w_{tj}^{(k)} \log(\alpha) + p(\alpha) \right) \\
\zeta_{1j}^{(k+1)} &= \frac{\sum_{t=1}^n (1 - w_{tj}^{(k)})(X_t - \sum_{\tau=1}^p \phi_{\tau j}^{(k)} X_{t-\tau})}{\sum_{t=1}^n (1 - w_{tj}^{(k)})} \\
\zeta_{2j}^{(k+1)} &= \frac{\sum_{t=1}^n w_{tj}^{(k)} (X_t - \sum_{\tau=1}^p \phi_{\tau j}^{(k)} X_{t-\tau})}{\sum_{t=1}^n w_{tj}^{(k)}} \\
\boldsymbol{\phi}_j^{(k+1)} &= \arg\max_{\boldsymbol{\phi}} \left(\sum_{t=1}^n (1 - w_{tj}^{(k)}) \log g(X_t|X_{t-1}^p; \zeta_{1j}^{(k+1)}, \boldsymbol{\phi}, \sigma_{1j}^{(k)}) \right. \\
&\quad \left. + \sum_{t=1}^n w_{tj}^{(k)} \log g(X_t|X_{t-1}^p; \zeta_{2j}^{(k+1)}, \boldsymbol{\phi}, \sigma_{2j}^{(k)}) \right) \\
\sigma_{1j}^{(k+1)} &= \arg\max_{\sigma_1} \sum_{t=1}^n (1 - w_{tj}^{(k)}) \log g(X_t|X_{t-1}^p; \zeta_{1j}^{(k+1)}, \boldsymbol{\phi}_j^{(k+1)}, \sigma_1) \\
\sigma_{2j}^{(k+1)} &= \arg\max_{\sigma_2} \sum_{t=1}^n w_{tj}^{(k)} \log g(X_t|X_{t-1}^p; \zeta_{2j}^{(k+1)}, \boldsymbol{\phi}_j^{(k+1)}, \sigma_2).
\end{aligned}$$

Compute

$$\begin{aligned}
M_n^{(k+1)}(\alpha_j) = 2 \Big\{ & pl_n(\alpha_j^{(k+1)}, \zeta_{1j}^{(k+1)}, \zeta_{2j}^{(k+1)}, \boldsymbol{\phi}_j^{(k+1)}, \sigma_{1j}^{(k+1)}, \sigma_{2j}^{(k+1)}) \\
& - pl_n(0.5, \widehat{\zeta}_0, \widehat{\zeta}_0, \widehat{\boldsymbol{\phi}}_0, \widehat{\sigma}_0, \widehat{\sigma}_0) \Big\},
\end{aligned}$$

put $k = k + 1$ and repeat Step 3 for a fixed number of iterations K.

Step 4. Put $j = j + 1$, $k = 0$ and go to Step 1, until $j = J$.

Step 5. Compute the test statistic

$$EM_n^{(K)} = \max\left\{M_n^{(K)}(\alpha_j), j = 1, \ldots, J\right\}.$$

4.3 Asymptotics

In this section, we give the asymptotic distribution of the (quasi) EM-test for testing the hypothesis of one regime, i.e. $\mathcal{M} = \{1\}$ against the alternative of (at least) two regimes, i.e. $\mathcal{M} = \{1, 2\}$, in model (4.1.1). The following theorem shows that the previously introduced EM-test admits the same asymptotic distribution as the EM-test, introduced by Chen and Li (2009), for testing for homogeneity in a normal mixture model with possibly distinct means and variances under the alternative.

Theorem 4.1. *Let $p(\alpha)$ be a continuous function, such that $p(\alpha) \to -\infty$ as $\alpha \to 0$ that attains its maximal value at $1/2$. Under the null model and for every fixed K, we have whenever one of the initial values α_j is equal to $1/2$,*

$$EM_n^{(K)} \xrightarrow{d} \chi_2^2,\ n \to \infty,$$

where χ_2^2 denotes the χ^2 distribution with 2 degrees of freedom.

We defer the proof of this theorem to Section 4.6.

The proof of Theorem 4.1 shows, that the asymptotic distribution is dominated by $\alpha = 1/2$, in other words, the same asymptotic distribution arises if under the alternative $\alpha = 1/2$ is fixed. Therefore, we propose a test based on fixed proportion $\alpha = 1/2$ under the alternative. To this end, we define

$$R_n(\alpha_0) = 2\{l_n(\alpha_0, \widehat{\zeta}_{1,\alpha_0}, \widehat{\zeta}_{2,\alpha_0}, \widehat{\boldsymbol{\phi}}_{\alpha_0}, \widehat{\sigma}_{1,\alpha_0}, \widehat{\sigma}_{2,\alpha_0}) - l_n(0.5, \widehat{\zeta}_0, \widehat{\zeta}_0, \widehat{\boldsymbol{\phi}}_0, \widehat{\sigma}_0, \widehat{\sigma}_0)\},$$

where $(\widehat{\zeta}_{1,\alpha_0}, \widehat{\zeta}_{2,\alpha_0}, \widehat{\boldsymbol{\phi}}_{\alpha_0}, \widehat{\sigma}_{1,\alpha_0}, \widehat{\sigma}_{2,\alpha_0})$ is the maximizer of $l_n(\alpha, \zeta_1, \zeta_2, \boldsymbol{\phi}, \sigma_1, \sigma_2)$ subject to $\alpha = \alpha_0$, $\alpha_0 \in (0, 1)$. The following corollary shows that the asymptotic distribution of $R_n(1/2)$ is the same as for the corresponding EM-test.

Corollary 4.2. *Under the null model, we have*

$$R_n(1/2) \xrightarrow{d} \chi_2^2,\ n \to \infty.$$

Clearly, $R_n(1/2) \leq EM_n^{(K)}$ and therefore the χ_2^2 distribution is an asymptotic upper bound for $R_n(1/2)$. Choosing the same values $\tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\boldsymbol{\phi}}, \tilde{\sigma}_1$ and $\tilde{\sigma}_2$ as in the end of the proof of Theorem 4.1 it is also clear that the χ_2^2 distribution serves as an asymptotic lower bound for $R_n(1/2)$.

Remark 4.1. As in Chapter 3 we can define a test based on a set of fixed proportions, say $\mathcal{J} = \{\alpha_1, \ldots, \alpha_J\}$. The corresponding test statistic is given by $R_n(\mathcal{J}) = \max_{\alpha_j \in \mathcal{J}} R_n(\alpha_j)$. Even if $R_n(\mathcal{J}) \leq EM_n^{(K)}$ is not true in general, the proof of Theorem 4.1 shows that the χ_2^2 distribution serves as an asymptotic upper bound for $R_n(\mathcal{J})$. If $1/2 \in \mathcal{J}$ and using the same values as in the end of the proof of Theorem 4.1 it is also clear that this stochastic upper bound is also attained. Therefore, we have $R_n(\mathcal{J}) \xrightarrow{d} \chi_2^2$, $n \to \infty$, under the null model. But simulations indicate that this test is highly anticonservative for finite sample sizes. Using simulated critical values this test does not show higher power, compared to the test based on fixed $\alpha = 1/2$ under the alternative, for finite sample sizes.

4.4 Simulations

In this section, we present some of the results of an extensive simulation study of the EM-test.

4.4.1 Simulated sizes

In the following we simulate the size of the EM-test and of the test based on the fixed proportion $\alpha = 1/2$ under the alternative for some data generating processes (DGP). For the EM-test, we choose $\mathcal{J} = \{0.1, 0.3, 0.5\}$ and $p(\alpha) = C \log(1 - |1 - 2\alpha|)$, $C = 1, 3$. For the computation, we choose $\delta = 0.1$.
DGP 1: $X_t = 0.5X_{t-1} + \epsilon_t$ where $\epsilon_t \overset{iid}{\sim} N(0, 1)$.
Model 1: $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0, 1)$.
The results for various sample sizes are displayed in Table 4.1.

Table 4.1: DGP: $X_t = 0.5X_{t-1} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0, 1)$, Model: $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, with $\epsilon_t \overset{iid}{\sim} N(0, 1)$; number of replications: 20,000.

		$C = 1$			$C = 3$			
Sample Size	Levels (%)	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$R_n(1/2)$
$n = 200$	10%	18.6	19.2	19.4	11.8	12.0	12.0	11.1
	5%	10.7	11.2	11.3	6.4	6.4	6.4	5.8
	1%	2.9	3.1	3.1	1.4	1.4	1.4	1.2
$n = 500$	10%	13.8	14.1	14.9	10.9	10.9	10.9	10.7
	5%	7.6	7.9	8.1	5.6	5.6	5.6	5.4
	1%	2.1	2.1	2.2	1.4	1.4	1.4	1.3
$n = 1000$	10%	12.1	12.1	12.2	10.4	10.4	10.4	10.4
	5%	6.5	6.5	6.6	5.3	5.3	5.3	5.2
	1%	1.6	1.6	1.6	1.1	1.1	1.1	1.1

DGP 2: $X_t = 0.6X_{t-1} - 0.3X_{t-2} + \epsilon_t$ where $\epsilon_t \overset{iid}{\sim} N(0, 1)$.
Model 2: $X_t = \zeta_{S_t} + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \sigma_{S_t}\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0, 1)$.
The results for various sample sizes can be found in Table 4.2.

Table 4.2: DGP: $X_t = 0.6X_{t-1} - 0.3X_{t-2} + \epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, Model: $X_t = \zeta_{S_t} + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \sigma_{S_t}\epsilon_t$, with $\epsilon_t \overset{iid}{\sim} N(0,1)$; number of replications: 20,000.

		$C=1$			$C=3$			
Sample Size	Levels (%)	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$EM_n^{(0)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$R_n(1/2)$
$n=200$	10%	22.4	23.0	23.2	12.5	12.8	13.0	11.6
	5%	13.6	14.2	14.4	6.7	6.9	7.0	6.0
	1%	3.8	4.1	4.2	1.7	1.8	1.8	1.5
$n=500$	10%	14.4	14.7	14.9	11.1	11.1	11.1	10.8
	5%	8.2	8.5	8.6	5.9	5.9	5.9	5.7
	1%	2.2	2.3	2.4	1.2	1.2	1.2	1.1
$n=1000$	10%	12.5	12.6	12.7	10.8	10.8	10.8	10.7
	5%	6.7	6.8	6.9	5.4	5.4	5.4	5.3
	1%	1.5	1.5	1.5	1.0	1.0	1.0	1.0

Both tests are anticonservative for small sample sizes in both scenarios. The EM-test becomes much more conservative when increasing the constant C.

4.4.2 Power comparison of several tests

Here we conduct a power comparison between several tests, e.g. the EM-test ($EM_n^{(K)}$) and the test based on fixed $\alpha = 1/2$ under the alternative. Even if we do not know the asymptotic properties of the quasi likelihood ratio test (abbreviated QLRT), introduced by Cho and White (2007), in the case of a bivariate switching parameter, we use it as a benchmark test for our EM-test as well as for the test based on fixed $\alpha = 1/2$ under the alternative. Note here, that the QLRT statistic is given by

$$QLR_n = 2\{\max_{\alpha,\zeta_1,\zeta_2,\boldsymbol{\phi},\sigma_1,\sigma_2} l_n(\alpha,\zeta_1,\zeta_2,\boldsymbol{\phi},\sigma_1,\sigma_2) - l_n(0.5,\widehat{\zeta}_0,\widehat{\zeta}_0,\widehat{\boldsymbol{\phi}}_0,\widehat{\sigma}_0,\widehat{\sigma}_0)\}.$$

Clearly, χ_2^2 serves as an asymptotic lower bound for the QLRT. Applying this lower bound leads to an anticonservative test which is not desirable, in general. Therefore we use simulated critical values for all scenarios and all tests.
We compare the power of the proposed tests with the power of the EM-test designed for linear switching autoregressive models with possibly switching intercept under the alternative (cf. Chapter 3). We denote the corresponding test statistic by $\widetilde{EM}_n^{(K)}$, here. We choose the penalty function $p(\alpha) = \log(1 - |1-2\alpha|)$ and $\mathcal{J} = \{0.1, 0.3, 0.5\}$ for $EM_n^{(K)}$ and $\widetilde{EM}_n^{K}$.

DGP 1: $X_t = (-1)^{S_t} \cdot 0.1 + 0.5X_{t-1} + (\mathbb{1}_{\{S_t=1\}} + 2 \cdot \mathbb{1}_{\{S_t=2\}})\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$ and various combinations a_{12} and a_{21} leading to different values of α.
Model 1(a): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $EM_n^{(K)}$, QLR_n and $R_n(1/2)$ and Model 1(b): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $\widetilde{EM}_n^{(K)}$.
The results are presented in Table 4.3. Here, the power of the compared tests mainly depends on the stationary distribution of $(S_k)_k$. The particular transition probabilities of the hidden Markov chain do not significantly influence the power of the corresponding tests. In all scenarios, the test based on fixed $\alpha = 1/2$ under the alternative outperforms the EM-test. The EM-test designed for linear switching autoregressive models with possibly switching intercept under the alternative shows the lowest power of the tests under consideration. Therefore, we should use this test only if there is a priori knowledge that the scale parameter in both regimes is almost identical.

Table 4.3: Nominal level: 5% DGP: $X_t = (-1)^{S_t} \cdot 0.1 + 0.5X_{t-1} + (\mathbb{1}_{\{S_t=1\}} + 2 \cdot \mathbb{1}_{\{S_t=2\}})\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$, sample size: 200, number of replications: 5,000. Model (a): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $EM_n^{(K)}$, QLR_n and $R_n(1/2)$ and Model (b): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $\widetilde{EM}_n^{(K)}$. Let $\alpha = a_{12}/(a_{12} + a_{21})$ be the stationary distribution of the hidden Markov chain $(S_k)_k$.

a_{12}	a_{21}	α	$EM_n^{(0)}$	$EM_n^{(1)}$	$\widetilde{EM}_n^{(0)}$	$\widetilde{EM}_n^{(1)}$	QLR_n	$R_n(1/2)$
0.5	0.5	0.5	56.2	55.6	14.4	15.1	46.7	63.1
0.3	0.3		55.3	54.3	15.9	16.5	41.9	63.5
0.1	0.1		55.5	54.5	15.1	15.8	42.7	60.9
0.9	0.9		55.4	54.5	14.6	15.4	40.9	64.4
0.4	0.6	0.4	64.6	63.6	19.2	20.5	56.3	72.5
0.3	0.45		62.9	62.0	18.3	19.5	54.7	72.6
0.2	0.3		63.3	62.6	18.5	19.7	51.9	71.2
0.1	0.15		61.3	60.5	18.8	20.0	54.1	69.1
0.3	0.7	0.3	65.6	65.1	20.6	22.6	55.4	73.2
0.15	0.35		65.6	65.2	19.0	20.9	56.9	71.5
0.2	0.8	0.2	59.6	59.2	18.6	20.9	52.4	66.2
0.15	0.6		58.0	57.6	17.1	19.0	52.4	64.3
0.1	0.4		58.0	57.5	18.1	19.6	51.9	64.3
0.05	0.2		55.2	54.3	16.6	18.4	51.7	61.8
0.1	0.9	0.1	38.1	37.8	10.4	11.7	36.1	43.5
0.05	0.45		36.4	35.8	9.0	10.2	36.2	40.9

DGP 2: $X_t = (-1)^{S_t} + 0.3X_{t-1} + (\mathbb{1}_{\{S_t=1\}} + 1.1 \cdot \mathbb{1}_{\{S_t=2\}})\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$ and various combinations a_{12} and a_{21} leading to different values of α.
Model 2(a): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $EM_n^{(K)}$, QLR_n and $R_n(1/2)$ and Model 2(b): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$, where $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $\widetilde{EM}_n^{(K)}$.

The results are presented in Table 4.4. Here, the power of our tests depends to a great extent on the stationary distribution as well as on the transition probabilities of the hidden Markov chain. Since the scale parameter of the innovation process is quite close in both regimes, the EM-test designed for linear switching autoregressive models with possibly switching intercept under the alternative outperforms the tests which allow for a switch in the interept and in the variance.

Table 4.4: Nominal level: 5% DGP: $X_t = (-1)^{S_t} + 0.3X_{t-1} + (\mathbb{1}_{\{S_t=1\}} + 1.1 \cdot \mathbb{1}_{\{S_t=2\}})\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$, sample size: 200, number of replications: 5,000. Model (a): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $EM_n^{(K)}$, QLR_n and $R_n(1/2)$ and Model (b): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $\widetilde{EM}_n^{(K)}$. Let $\alpha = a_{12}/(a_{12} + a_{21})$ be the stationary distribution of the hidden Markov chain $(S_k)_k$.

a_{12}	a_{21}	α	$EM_n^{(0)}$	$EM_n^{(1)}$	$\widetilde{EM}_n^{(0)}$	$\widetilde{EM}_n^{(1)}$	QLR_n	$R_n(1/2)$
0.5	0.5	0.5	31.8	31.4	53.5	53.0	22.0	38.7
0.3	0.3		21.6	21.3	38.6	38.2	17.9	27.8
0.1	0.1		6.9	6.7	9.1	9.1	6.0	8.7
0.7	0.7		19.9	19.5	38.8	38.3	15.6	25.0
0.9	0.9		7.7	7.5	12.1	12.4	6.2	8.8
0.4	0.6	0.4	41.6	40.7	54.7	55.4	29.5	50.0
0.2	0.3		24.2	23.8	35.5	35.5	17.6	28.3
0.1	0.15		11.4	11.1	13.1	13.4	6.2	14.0
0.3	0.7	0.3	53.2	52.4	68.6	68.0	40.6	61.6
0.15	0.35		32.6	32.0	43.7	43.3	24.1	40.1
0.2	0.8	0.2	57.7	56.8	62.1	62.5	47.7	66.0
0.15	0.6		52.9	52.3	61.0	60.9	41.9	62.2
0.1	0.4		38.3	37.4	45.7	45.5	29.6	47.4
0.05	0.2		18.5	18.1	18.7	18.9	16.8	24.2
0.1	0.9	0.1	42.8	42.0	44.7	44.2	36.4	50.3
0.05	0.45		29.8	28.8	32.8	32.5	26.4	37.4

Finally, we compare the power of the tests under consideration with the power of the MQLRT designed for linear switching autoregressive models with possibly switching scale parameter of the innovations under the alternative (which has been introduced in Chapter 2). In the following, the corresponding test statistic will be denoted be M_n.

DGP 3: $X_t = (-1)^{S_t} \cdot 0.7 + 0.5X_{t-1} + (1.8 \cdot \mathbb{1}_{\{S_t=1\}} + \cdot\mathbb{1}_{\{S_t=2\}})\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$ and various combinations a_{12} and a_{21} leading to different values of α.
Model 3(a): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $EM_n^{(K)}$, QLR_n and $R_n(1/2)$, Model 3(b): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $\widetilde{EM}_n^{(K)}$, and Model 3(c): $X_t = \zeta + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, for M_n.
The results can be found in Table 4.5. Here, the tests designed for switching intercept

and variance under the alternative outperform $\widetilde{EM}_n^{(K)}$ as well as M_n. In all scenarios, the test based on fixed $\alpha = 1/2$ has higher power than the corresponding EM-test. Since M_n and $\widetilde{EM}_n^{(K)}$ have lower power than the tests designed for possibly switching intercept and variance under the alternative, $EM_n^{(K)}$ or $R_n(1/2)$ should be preferred if there is no a priori knowledge that the scale parameter (resp. the intercept) in both regimes is almost identical.

Table 4.5: Nominal level: 5% DGP: $X_t = (-1)^{S_t} \cdot 0.7 + 0.5X_{t-1} + (1.8 \cdot \mathbb{1}_{\{S_t=1\}} + \cdot \mathbb{1}_{\{S_t=2\}})\epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$, sample size: 200, number of replications: 5,000. Model (a): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $EM_n^{(K)}$, QLR_n and $R_n(1/2)$, Model (b): $X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, for $\widetilde{EM}_n^{(K)}$, and Model (c): $X_t = \zeta + \phi X_{t-1} + \sigma_{S_t}\epsilon_t$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, for M_n. Let $\alpha = a_{12}/(a_{12}+a_{21})$ be the stationary distribution of the hidden Markov chain $(S_k)_k$.

a_{12}	a_{21}	α	$EM_n^{(0)}$	$EM_n^{(1)}$	$\widetilde{EM}_n^{(0)}$	$\widetilde{EM}_n^{(1)}$	M_n	QLR_n	$R_n(1/2)$
0.5	0.5	0.5	82.2	81.5	47.4	51.3	36.6	70.5	87.7
0.3	0.3		78.6	77.9	54.2	59.6	50.2	67.0	85.4
0.1	0.1		69.0	68.0	35.6	41.3	48.8	57.8	75.3
0.7	0.7		80.3	79.9	51.4	55.4	39.7	69.6	86.3
0.9	0.9		76.9	76.1	47.1	51.9	38.5	65.1	84.3
0.4	0.6	0.4	66.4	65.6	49.7	52.2	18.9	52.9	75.2
0.2	0.3		63.7	62.6	42.4	45.4	22.8	50.6	71.7
0.2	0.8	0.2	21.1	20.6	23.3	23.6	4.3	15.3	26.7
0.1	0.4		22.8	22.6	21.0	22.0	5.0	16.9	26.6
0.05	0.2		21.6	21.4	16.8	17.7	7.1	16.5	27.2

4.5 Application

In this section, we apply our methods to the series of monthly log returns $(X_t)_{t=-3,\dots,884}$ of IBM stock from January 1926 to December 1999. The returns are in percentage and include dividends. The data can be obtained from

http://faculty.chicagobooth.edu/ruey.tsay/teaching/fts/m-ibmspln.dat.

Marginal distribution

The histogram and a kernel density estimate (see Figure 4.1) indicate slight asymmetry and fat tails. The empirical skewness coefficient and kurtosis are given by -0.2369 and 4.9278, respectively. To deal with skewness and kurtosis in the unconditional distribution of stock returns finite normal mixtures have been applied quite often, see e.g. Kon (1984).

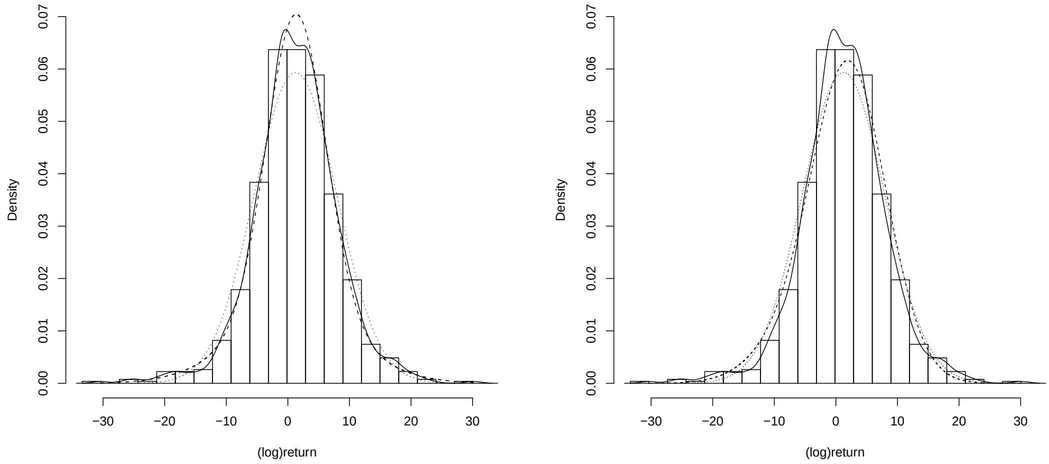

Figure 4.1: Histogram together with the density of a fitted normal distribution (dotted line), a fitted two-component normal mixture with distinct means and variances (dashed line) and a kernel density estimate (solid line) of monthly log returns for IBM stock (left) and together with the density of a fitted normal distribution (dotted line), a fitted two-component homoscedastic normal mixture (dashed line) and a kernel density estimate (solid line) of monthly log returns for IBM stock (right).

In a first step, we test one against two components in a normal mixture model using penalized likelihood based tests. The hypothesis of a single component is rejected (with p-value < 0.001) by every test under consideration. Testing against an alternative with possibly distinct means and variances using the EM-test introduced in Chen and Li (2009), we see that the alternative two-component model has almost identical means ($\overline{\mu}_1 = 1.37$ and $\overline{\mu}_2 = 0.94$), quite different standard deviations ($\overline{\sigma}_1 = 4.80$ and $\overline{\sigma}_2 = 9.89$) and the relative size of component 2 $\overline{\alpha}$ is 0.30.
But modeling the series of log returns for IBM stock by finite mixtures would only be appropriate if (X_t) did not exhibit autocorrelation. As can be seen in Figure 4.2 (right) this is not the case for our time series and therefore we model (X_t) by autoregressive models.

Autoregressive model

As indicated by the pacf, see Figure 4.2 (right), we fit an AR(1) model

$$X_t = \zeta_0 + \phi_0 X_{t-1} + \sigma_0 \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0,1),$$

yielding the estimate $(\widehat{\zeta}_0, \widehat{\phi}_0, \widehat{\sigma}_0) = (1.157, 0.077, 6.698)$, to capture autocorrelation in our time series. Computing the residuals $(\widehat{\epsilon}_t)$ of the fitted model and testing of normality using Anderson-Darling test ($A_n = 2.95$) we strongly reject $H : \epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$ by a p-value < 0.001 (why to use asymptotic critical values of the Anderson-Darling test for independent

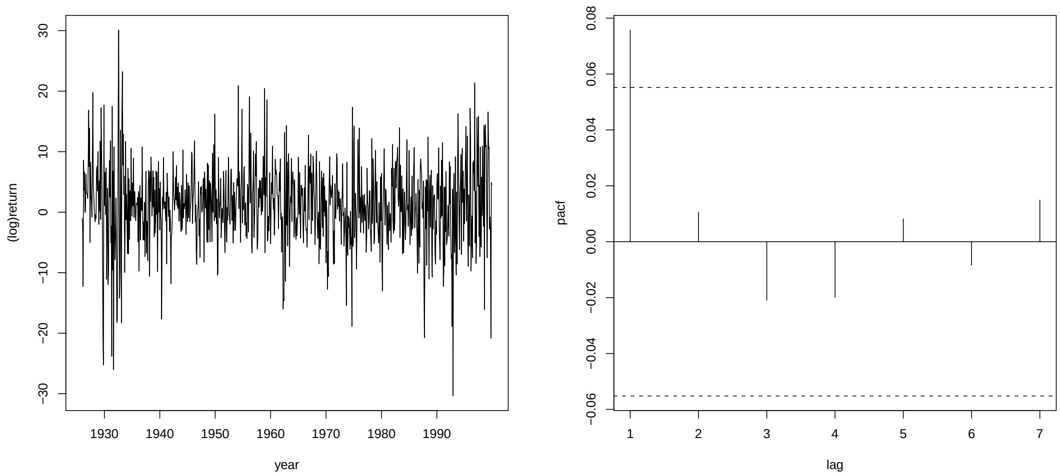

Figure 4.2: Monthly log returns (in % and including dividends) for IBM stock from January 1926 to December 1999 (left). Sample partial autocorrelation function of monthly log returns for IBM stock. The dashed line gives an approximate pointwise 90% confidence interval (right).

and identically distributed observations, cf. Pierce 1985) and which indicates lack-of-fit of the supposed AR(1) model.
While Tsay (2002) fits an AR(1)-GARCH(1,1) to the monthly log returns of IBM stock Kim, Nelson and Startz (1998) and Bhar and Hamori (2004) suggest modeling monthly stock returns by Markov-switching autoregressive models, in general. We follow this approach and fit several linear switching autoregressive models to the data.

First, we test the hypothesis of one regime against the alternative of possibly switching intercept and variance, i.e.

$$X_t = \zeta_{S_t} + \phi X_{t-1} + \sigma_{S_t}\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0,1), \tag{4.5.1}$$

using the EM-test introduced in this chapter. We find $EM_n^{(2)} = 54.54$ and can reject the hypothesis of one regime by a p-value < 0.001. The corresponding EM-estimate $(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}_1, \overline{\sigma}_2)$ is $(0.248, 1.293, 0.926, 0.041, 4.948, 10.352)$. The full model MLE $\hat{\boldsymbol{\omega}}$ (conditional on $X_0 = x_0$ and $S_0 = 1$) in model (4.5.1) yields $\hat{\boldsymbol{\omega}} = (\widehat{a}_{12}, \widehat{a}_{21}, \widehat{\zeta}_1, \widehat{\zeta}_2, \widehat{\phi}, \widehat{\sigma}_1, \widehat{\sigma}_2) = (0.015, 0.052, 1.266, 0.752, 0.081, 5.183, 10.383)$. Our analysis indicates that there are two regimes: Regime 1 with higher mean level in the (log) returns and lower variance and regime 2 with lower mean level in the (log) returns and higher variance. Comparing several (switching) autoregressive models

$$\mathcal{M}_1: \quad X_t \;=\; \zeta + \sum_{j=1}^{p} \phi_j X_{t-j} + \sigma\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0,1),$$

$$\begin{aligned}
\mathcal{M}_2: \quad X_t &= \zeta_{S_t} + \sum_{j=1}^{p} \phi_{j,S_t} X_{t-j} + \sigma_{S_t}\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0,1), \\
\mathcal{M}_3: \quad X_t &= \zeta + \sum_{j=1}^{p} \phi_j X_{t-j} + \sigma_{S_t}\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0,1), \\
\mathcal{M}_4: \quad X_t &= \zeta_{S_t} + \sum_{j=1}^{p} \phi_j X_{t-j} + \sigma\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0,1), \\
\mathcal{M}_5: \quad X_t &= \zeta_{S_t} + \sum_{j=1}^{p} \phi_j X_{t-j} + \sigma_{S_t}\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0,1),
\end{aligned}$$

where S_t takes values in $\mathcal{M} = \{1, 2\}$ and $p = 1, \ldots, 4$, based on formal model selection criteria such as BIC or AIC one chooses model $\mathcal{M}_3$ with $p = 1$, see Table 4.6 and 4.7. Here, we note that the AIC and BIC are computed by

$$\text{AIC} = -2\tilde{l}_n(\hat{\boldsymbol{\omega}}) + 2 \cdot k(\hat{\boldsymbol{\omega}}) \text{ and } \text{BIC} = -2\tilde{l}_n(\hat{\boldsymbol{\omega}}) + \log(n) \cdot k(\hat{\boldsymbol{\omega}}),$$

where $\tilde{l}_n(\cdot)$ is the full model log likelihood conditional on the first 4 observations and on state 1 and $k(\hat{\boldsymbol{\omega}})$ denotes the length of $\hat{\boldsymbol{\omega}}$.

Table 4.6: BIC for the corresponding models for monthly returns of IBM stock

BIC	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$
1	5891.56	5812.57	**5799.38**	5911.92	5805.84
2	5898.25	5825.86	5805.96	5918.60	5812.42
3	5904.65	5838.91	5812.71	5925.00	5819.17
4	5911.08	5849.97	5817.83	5931.43	5824.29

Table 4.7: AIC for the corresponding models for monthly returns of IBM stock

AIC	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$
1	5877.21	5774.30	**5770.67**	5883.21	5772.35
2	5879.11	5778.01	5772.47	5885.11	5774.15
3	5880.73	5781.50	5774.43	5886.73	5776.11
4	5882.37	5782.99	5774.77	5888.37	5776.45

Testing $H : \zeta_1 = \zeta_2$ in model $\mathcal{M}_5$ is a regular problem and therefore the likelihood ratio statistic

$$T_n = 2\big\{\max_{a_{12},a_{21},\zeta_1,\zeta_2,\phi,\sigma_1,\sigma_2} \tilde{l}_n(a_{12}, a_{21}, \zeta_1, \zeta_2, \phi, \sigma_1, \sigma_2) - \max_{a_{12},a_{21},\zeta,\phi,\sigma_1,\sigma_2} \tilde{l}_n(a_{12}, a_{21}, \zeta, \zeta, \phi, \sigma_1, \sigma_2)\big\}$$

asymptotically follows a χ_1^2 distribution. In our case, we have $T_n = 0.322$ (p-value=0.57). Therefore, we cannot reject the hypothesis $H : \zeta_1 = \zeta_2$. Testing the hypothesis $H : \sigma_1 = \sigma_2$ in model $\mathcal{M}_5$ is also a regular problem and therefore the likelihood ratio statistic

$$T_n = 2\big\{ \max_{a_{12},a_{21},\zeta_1,\zeta_2,\phi,\sigma_1,\sigma_2} \tilde{l}_n(a_{12}, a_{21}, \zeta_1, \zeta_2, \phi, \sigma_1, \sigma_2) - \max_{a_{12},a_{21},\zeta_1,\zeta_2,\phi,\sigma} \tilde{l}_n(a_{12}, a_{21}, \zeta_1, \zeta_2, \phi, \sigma, \sigma) \big\}$$

asymptotically follows a χ_1^2 distribution. In our case, we have $T_n = 112.8643$ (p-value < 0.001). Therefore, we clearly reject the hypothesis $H : \sigma_1 = \sigma_2$. These results motivate testing for homogeneity in the model

$$X_t = \zeta + \phi X_{t-1} + \sigma_{S_t}\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0,1), \tag{4.5.2}$$

where the intercept ζ is treated as a structural parameter, i.e. it is equal in every state, using the MQLRT introduced in Chapter 2. In this case, we can reject the hypothesis of one regime by a p-value < 0.001 ($M_n = 54.41$). The modified MLE $(\widehat{\alpha}^*, \widehat{\zeta}^*, \widehat{\phi}^*, \widehat{\sigma}_1^*, \widehat{\sigma}_2^*)$ is given by $(0.248, 1.234, 0.042, 4.943, 10.354)$. The full model conditional MLE $\hat{\boldsymbol{\omega}}$ in model (4.5.2) yields $\widehat{\boldsymbol{\omega}} = (\widehat{a}_{12}, \widehat{a}_{21}, \widehat{\zeta}, \widehat{\phi}, \widehat{\sigma}_1, \widehat{\sigma}_2) = (0.016, 0.053, 1.210, 0.081, 5.164, 10.371)$. Therefore, model (4.5.1) as well as (4.5.2) seem to be appropriate models for the series of log returns for IBM stock.

4.6 Proofs

To prove Theorem 4.1, we need some additional lemmas. Since we assume that the innovations $(\epsilon_k)_k$ are independent and identically normally distributed the conditional density (w.r.t. Lebesgue measure on $\mathbb{R}$) of X_t given $X_{t-1}^p = x_{t-1}^p$ and $S_t = i$ is

$$g(x_t|x_{t-1}^p; \zeta_i, \boldsymbol{\phi}, \sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{x_t - \zeta_i - \sum_{j=1}^p \phi_j x_{t-j}}{2\sigma_i^2} \right).$$

In the following, let $(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\boldsymbol{\phi}}, \overline{\sigma}_1, \overline{\sigma}_2)$ be any EM-estimator.

Lemma 4.1. *For each given $\overline{\alpha} \in (0, 0.5]$ we have under the null model*

$$\begin{aligned} \overline{\sigma}_1 - \sigma_0 &= o_P(1), & \overline{\sigma}_2 - \sigma_0 &= o_P(1), \\ \overline{\zeta}_1 - \zeta_0 &= o_P(1), & \overline{\zeta}_2 - \zeta_0 &= o_P(1), \\ \overline{\boldsymbol{\phi}} - \boldsymbol{\phi}_0 &= o_P(1). \end{aligned}$$

The proof of Lemma 4.1 follows the lines of the proof Lemma 3.1 and is therefore omitted.

From now on, we confine our attention to linear autoregressive models of order 1. The extension to autoregressive models of higher order, i.e. $p > 1$, will be analogous to the case of the variance being a structural parameter instead of a switching parameter. Without loss of generality we assume $\zeta_0 = 0$ and $\sigma_0 = 1$.

Note that we can assume that $(\overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}_1, \overline{\sigma}_2)$ are in a small neighorhood of $(0, 0, \phi_0, 1, 1)$ by Lemma 4.1 whenever $\delta' < \overline{\alpha} < 1 - \delta'$ for any $\delta' \in (0, 0.5]$. As in Chapter 3 we define

$$\begin{aligned} H_n(\alpha) &= \left(n - \sum_{t=1}^{n} \overline{w}_t\right) \log(1-\alpha) + \sum_{t=1}^{n} \overline{w}_t \log(\alpha) + p(\alpha) \\ &=: R_n(\alpha) + p(\alpha), \end{aligned}$$

where

$$\overline{w}_t = \frac{\overline{\alpha} g(X_t | X_{t-1}; \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}_2)}{(1-\overline{\alpha}) g(X_t | X_{t-1}; \overline{\zeta}_1, \overline{\phi}, \overline{\sigma}_1) + \overline{\alpha} g(X_t | X_{t-1}; \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}_2)}.$$

Let $\overline{\alpha}^* = \arg\max_{\alpha \in [0,1]} H_n(\alpha)$. The following lemma shows that the EM-iteration changes the fitted value of α by no more than $o_p(1)$.

Lemma 4.2. *Under the conditions of Lemma 4.1 and if $\overline{\alpha} - \alpha_j = o_p(1)$ for some $\alpha_j \in (0, 1)$, then we have under the null model*

$$\overline{\alpha}^* - \alpha_j = o_P(1).$$

The proof of Lemma 4.2 is essentially the same as the proof of Lemma 3.4 and is therefore omitted.

In a first step give a stochastic upper bound for

$$2\{pl_n(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}_1, \overline{\sigma}_2) - pl_n(0.5, \widehat{\zeta}_0, \widehat{\zeta}_0, \widehat{\phi}_0, \widehat{\sigma}_0, \widehat{\sigma}_0)\}. \tag{4.6.1}$$

Note that

$$\begin{aligned} &2\{pl_n(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}_1, \overline{\sigma}_2) - pl_n(0.5, \widehat{\zeta}_0, \widehat{\zeta}_0, \widehat{\phi}_0, \widehat{\sigma}_0, \widehat{\sigma}_0)\} \\ =& 2\{l_n(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}_1, \overline{\sigma}_2) - l_n(0.5, \widehat{\zeta}_0, \widehat{\zeta}_0, \widehat{\phi}_0, \widehat{\sigma}_0, \widehat{\sigma}_0)\} \\ &+ 2\{p(\overline{\alpha}) - p(0.5)\} \\ \leq& r_{1n}(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}_1, \overline{\sigma}_2) + r_{2n}, \end{aligned} \tag{4.6.2}$$

where

$$\begin{aligned} r_{1n} &= r_{1n}(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}_1, \overline{\sigma}_2) = 2\{l_n(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}_1, \overline{\sigma}_2) - l_n(0.5, 0, 0, \phi_0, 1, 1)\}, \\ r_{2n} &= 2\{l_n(0.5, 0, 0, \phi_0, 1, 1) - l_n(0.5, \widehat{\zeta}_0, \widehat{\zeta}_0, \widehat{\phi}_0, \widehat{\sigma}_0, \widehat{\sigma}_0)\}. \end{aligned}$$

The last inequality in (4.6.2) follows from the properties of the penalty function $p(\alpha)$. Letting $p(\alpha) \equiv 0$, e.g. in the case of a test based on fixed proportions the inequality in (4.6.2) becomes an equality.

Let $r_{1n}(\overline{\alpha}, \overline{\zeta}_1, \overline{\zeta}_2, \overline{\phi}, \overline{\sigma}_1, \overline{\sigma}_2) = 2\sum_{t=1}^n \log(1+\overline{\delta}_t)$ with

$$\overline{\delta}_t = (1-\overline{\alpha})\left\{\frac{g(X_t|X_{t-1};\overline{\zeta}_1,\overline{\phi},\overline{\sigma}_1)}{g(X_t|X_{t-1};0,\phi_0,1)} - 1\right\} + \overline{\alpha}\left\{\frac{g(X_t|X_{t-1};\overline{\zeta}_2,\overline{\phi},\overline{\sigma}_2)}{g(X_t|X_{t-1};0,\phi_0,1)} - 1\right\}.$$

Using the inequality $\log(1+x) \le x - (1/2)x^2 + (1/3)x^3$ leads to

$$r_{1n} \le 2\sum_{t=1}^n \overline{\delta}_t - \sum_{t=1}^n \overline{\delta}_t^2 + (2/3)\sum_{t=1}^n \overline{\delta}_t^3. \tag{4.6.3}$$

For $0 \le l, s, i \le 4$ we define

$$\overline{m}_{l,s,i} = (1-\overline{\alpha})\overline{\zeta}_1^l(\overline{\sigma}_1^2-1)^s(\overline{\phi}-\phi_0)^i + \overline{\alpha}\overline{\zeta}_2^l(\overline{\sigma}_2^2-1)^s(\overline{\phi}-\phi_0)^i.$$

Denoting

$$\partial_\zeta^l \partial_{\sigma^2}^s \partial_\phi^i g(X_t|X_{t-1};0,\phi_0,1) = \left.\frac{\partial^{l+s+i} g(X_t|X_{t-1};\zeta,\phi,\sigma)}{\partial^l\zeta\,\partial^s(\sigma^2)\,\partial^i\phi}\right|_{(\zeta,\phi,\sigma)=(0,\phi_0,1)}$$

and expanding $g(X_t|X_{t-1};\overline{\zeta}_h,\overline{\phi},\overline{\sigma}_h)$, $h=1,2$, up to order 4, we get

$$\overline{\delta}_t = \sum_{l+s+i=1}^{4} \frac{1}{l!s!i!}\frac{\partial_\zeta^l \partial_{\sigma^2}^s \partial_\phi^i g(X_t|X_{t-1};0,\phi_0,1)}{g(X_t|X_{t-1};0,\phi_0,1)} + \overline{\epsilon}_{tn}^{(1)} \tag{4.6.4}$$

with remainder $\overline{\epsilon}_{tn}^{(1)}$ for which

$$\overline{\epsilon}_n^{(1)} := \sum_{t=1}^n \overline{\epsilon}_{tn}^{(1)} = O_P(n^{1/2})\left\{\sum_{h=1}^2 \sum_{\substack{l+s+i=5\\ l,s,i\ge 0}} |\overline{\zeta}_h|^l |\overline{\sigma}_h^2 - 1|^s |\overline{\phi}-\phi_0|^i\right\}. \tag{4.6.5}$$

This is due to the CLT for stationary and ergodic martingale differences.

We reexamine the remainder term $\overline{\epsilon}_n^{(1)}$ in order to simplify it. To this end we distinguish three cases:

Let $i=0$. In this case we have $l+s=5$. Following the assessment in Chen and Li (2009), we have by Lemma 4.1 for $l=0,1,2$

$$O_P(n^{1/2})\left\{\sum_{h=1}^2 |\overline{\zeta}_h|^l |\overline{\sigma}_h^2-1|^{5-l}\right\} = O_P(n^{1/2})\left\{\sum_{h=1}^2 |\overline{\sigma}_h^2-1|^3\right\}$$

and for $l = 3, 4$

$$O_P(n^{1/2}) \left\{ \sum_{h=1}^{2} |\overline{\zeta}_h|^l |\overline{\sigma}_h^2 - 1|^{5-l} \right\} = O_P(n^{1/2}) \left\{ \sum_{h=1}^{2} |\overline{\zeta}_h|^3 |\overline{\sigma}_h^2 - 1| \right\}.$$

For $i = 1$, we have $l + s = 4$. For $l = 0, 1$, we obtain

$$O_P(n^{1/2}) \left\{ \sum_{h=1}^{2} |\overline{\zeta}_h|^l |\overline{\sigma}_h^2 - 1|^{4-l} |\overline{\phi} - \phi_0| \right\} = O_P(n^{1/2}) \left\{ \sum_{h=1}^{2} |\overline{\sigma}_h^2 - 1|^3 \right\},$$

for $l = 2$

$$O_P(n^{1/2}) \left\{ \sum_{h=1}^{2} \overline{\zeta}_h^2 (\overline{\sigma}_h^2 - 1)^2 |\overline{\phi} - \phi_0| \right\} = O_P(n^{1/2}) \left\{ \sum_{h=1}^{2} |\overline{\zeta}_h| (\overline{\sigma}_h^2 - 1)^2 \right\},$$

for $l = 3$

$$O_P(n^{1/2}) \left\{ \sum_{h=1}^{2} |\overline{\zeta}_h|^3 |\overline{\sigma}_h^2 - 1| |\overline{\phi} - \phi_0| \right\} = O_P(n^{1/2}) \left\{ \sum_{h=1}^{2} \{ |\overline{\zeta}_h|^5 + (\overline{\phi} - \phi_0)^2 \} \right\}$$

which is due to

$$\begin{aligned} |\overline{\zeta}_h|^3 |\overline{\sigma}_h^2 - 1| |\overline{\phi} - \phi_0| &= |\overline{\zeta}_h| \overline{\zeta}_h^2 |\overline{\sigma}_h^2 - 1| |\overline{\phi} - \phi_0| \\ &\leq |\overline{\zeta}_h| (\overline{\zeta}_h^4 + |\overline{\sigma}_h^2 - 1|^2 (\overline{\phi} - \phi_0)^2) \\ &= |\overline{\zeta}_h|^5 + o_P(1)(\overline{\phi} - \phi_0)^2 \end{aligned}$$

by Lemma 4.1 and the inequality $ab \leq a^2 + b^2$. Finally, for $l = 4$, we have

$$O_P(n^{1/2}) \left\{ \sum_{h=1}^{2} \overline{\zeta}_h^4 |\overline{\phi} - \phi_0| \right\} = O_P(n^{1/2}) \left\{ \sum_{h=1}^{2} \{ |\overline{\zeta}_h|^5 + (\overline{\phi} - \phi_0)^2 \} \right\}$$

since

$$\begin{aligned} |\overline{\zeta}_h|^4 |\overline{\phi} - \phi| &\leq \overline{\zeta}_h^8 + (\overline{\phi} - \phi_0)^2 \\ &= o_P(1) |\overline{\zeta}_h|^5 + (\overline{\phi} - \phi)^2. \end{aligned}$$

If $i \geq 2$, we get

$$O_P(n^{1/2}) \left\{ \sum_{h=1}^{2} \sum_{\substack{l+s+i=5 \\ l,s,i \geq 0}} |\overline{\zeta}_h|^l |\overline{\sigma}_h^2 - 1|^s |\overline{\phi} - \phi_0|^i \right\} = O_P(n^{1/2})(\overline{\phi} - \phi_0)^2 \tag{4.6.6}$$

which is due to Lemma 4.1.

Therefore the remainder term $\overline{\epsilon}_n^{(1)}$ can be simplified to

$$\overline{\epsilon}_n^{(1)} = O_P(n^{1/2}) \sum_{h=1}^{2} \{|\overline{\zeta}_h|^5 + |\overline{\zeta}_h|^3 |\overline{\sigma}_h^2 - 1| + |\overline{\sigma}_h^2 - 1|^3 + (\overline{\phi} - \phi_0)^2\}. \tag{4.6.7}$$

By Lemma 4.1 we can incorporate the terms $\overline{m}_{l,s,i}$ with $l + 2s + 4i \geq 5$ into the remainder term, e.g.

$$\begin{aligned} O_P(n^{1/2})\overline{m}_{4,1,0} &= O_P(n^{1/2})\overline{\zeta}_h^4(\overline{\sigma}_h^2 - 1) = O_P(n^{1/2})\overline{\zeta}_h^3|\overline{\sigma}_h^2 - 1|, \\ O_P(n^{1/2})\overline{m}_{0,4,0} &= O_P(n^{1/2})(\overline{\sigma}_h^2 - 1)^4 = O_P(n^{1/2})|\overline{\sigma}_h^2 - 1|^3, \\ O_P(n^{1/2})\overline{m}_{2,2,0} &= O_P(n^{1/2})\overline{\zeta}_h^2(\overline{\sigma}_h^2 - 1)^2 = O_P(n^{1/2})|\overline{\zeta}_h|(\overline{\sigma}_h^2 - 1)^2 \end{aligned} \tag{4.6.8}$$

and

$$O_P(n^{1/2})\overline{m}_{i,j,2} = O_P(n^{1/2})|\overline{\zeta}_h|^i|\overline{\sigma}_h^2 - 1|^j(\overline{\phi} - \phi_0)^2 = O_P(n^{1/2})(\overline{\phi} - \phi_0)^2 \tag{4.6.9}$$

for $i, j \geq 0$. Adding

$$O_P(n^{1/2}) \sum_{h=1}^{2} \{|\overline{\zeta}_h||\overline{\phi} - \phi_0| + |\overline{\sigma}_h^2 - 1||\overline{\phi} - \phi_0|\}$$

to the remainder assures that the terms $\overline{m}_{l,s,1}$ with $l + 2s \geq 1$ can be subsumed into the remainder term.

Altogether, we get

$$\overline{\delta}_t = \sum_{l+2s+4i=1}^{4} \frac{1}{l!s!i!} \frac{\partial_\zeta^l \partial_{\sigma^2}^s \partial_\phi^i g(X_t|X_{t-1}; 0, \phi_0, 1)}{g(X_t|X_{t-1}; 0, \phi_0, 1)} + \overline{\epsilon}_{tn} \tag{4.6.10}$$

with remainder $\overline{\epsilon}_{tn}$ fulfilling

$$\begin{aligned} \overline{\epsilon}_n = \sum_{t=1}^{n} \overline{\epsilon}_{tn} &= O_P(n^{1/2}) \sum_{h=1}^{2} \{|\overline{\zeta}_h|^5 + |\overline{\zeta}_h|^3|\overline{\sigma}_h^2 - 1| + |\overline{\sigma}_h^2 - 1|^3 \\ &\quad + (\overline{\phi} - \phi_0)^2 + |\overline{\zeta}_h||\overline{\phi} - \phi_0| + |\overline{\sigma}_h^2 - 1||\overline{\phi} - \phi_0|\}. \end{aligned} \tag{4.6.11}$$

Using the inequality $ab \leq a^2 + b^2$, $a, b \in \mathbb{R}$, we see that

$$|\overline{\zeta}_h|^3|\overline{\sigma}_h^2 - 1| = |\overline{\zeta}_h|\overline{\zeta}_h^2|\overline{\sigma}_h^2 - 1| \leq |\overline{\zeta}_h|\{\overline{\zeta}_h^4 + (\overline{\sigma}_h^2 - 1)^2\} = |\overline{\zeta}_h|^5 + |\overline{\zeta}_h|(\overline{\sigma}_h^2 - 1)^2.$$

Therefore the remainder term can be written as

$$\bar{\epsilon}_n = \sum_{t=1}^{n} \bar{\epsilon}_{tn} = O_P(n^{1/2}) \sum_{h=1}^{2} \{|\bar{\zeta}_h|^5 + |\bar{\zeta}_h|(\bar{\sigma}_h^2 - 1)^2 + |\bar{\sigma}_h^2 - 1|^3 + (\bar{\phi} - \phi_0)^2 + |\bar{\zeta}_h||\bar{\phi} - \phi_0| + |\bar{\sigma}_h^2 - 1||\bar{\phi} - \phi_0|\}. \tag{4.6.12}$$

With Y_t, Z_t, U_t, V_t and W_{1t} defined as in the previous chapter we get

$$\bar{\delta}_t = \bar{t}_1 Y_t + \bar{t}_2 Z_t + \bar{t}_3 U_t + \bar{t}_4 V_t + \bar{t}_5 W_{1t} + \bar{\epsilon}_{tn} \tag{4.6.13}$$

with $\sum_{t=1}^{n} \bar{\epsilon}_{tn}$ satisfying (4.6.12) and

$$\begin{aligned} \bar{t}_1 &= \overline{m}_{1,0,0}, \\ \bar{t}_2 &= \overline{m}_{2,0,0} + \overline{m}_{0,1,0}, \\ \bar{t}_3 &= \overline{m}_{3,0,0} + 3\overline{m}_{1,1,0}, \\ \bar{t}_4 &= \overline{m}_{4,0,0} + 6\overline{m}_{2,1,0} + 3\overline{m}_{0,2,0}, \\ \bar{t}_5 &= \overline{m}_{0,0,1} = \bar{\phi} - \phi_0. \end{aligned} \tag{4.6.14}$$

Putting $\bar{\delta}_t$ into (4.6.3) and noting that the remainders from the square and cubic terms on the right-hand side of the following equation are of the same or higher order than the remainder $\bar{\epsilon}_n$ from the linear sum, we get

$$\begin{aligned} r_{1n} \le\, & 2\{\bar{t}_1 \sum_{t=1}^{n} Y_t + \bar{t}_2 \sum_{t=1}^{n} Z_t + \bar{t}_3 \sum_{t=1}^{n} U_t + \bar{t}_4 \sum_{t=1}^{n} V_t + \bar{t}_5 \sum_{t=1}^{n} W_{1t}\} \\ & - \{\bar{t}_1 \sum_{t=1}^{n} Y_t + \bar{t}_2 \sum_{t=1}^{n} Z_t + \bar{t}_3 \sum_{t=1}^{n} U_t + \bar{t}_4 \sum_{t=1}^{n} V_t + \bar{t}_5 \sum_{t=1}^{n} W_{1t}\}^2 \\ & + \frac{2}{3}\{\bar{t}_1 \sum_{t=1}^{n} Y_t + \bar{t}_2 \sum_{t=1}^{n} Z_t + \bar{t}_3 \sum_{t=1}^{n} U_t + \bar{t}_4 \sum_{t=1}^{n} V_t + \bar{t}_5 \sum_{t=1}^{n} W_{1t}\}^3 \\ & + O_P(\bar{\epsilon}_n). \end{aligned} \tag{4.6.15}$$

As in Chapter 3, one shows that

$$\left\{\bar{t}_1 \sum_{t=1}^{n} Y_t + \bar{t}_2 \sum_{t=1}^{n} Z_t + \bar{t}_3 \sum_{t=1}^{n} U_t + \bar{t}_4 \sum_{t=1}^{n} V_t + \bar{t}_5 \sum_{t=1}^{n} W_{1t}\right\}^3 = o_P(n) \left\{\sum_{l=1}^{5} \bar{t}_l^2\right\}.$$

Due to the non-degeneracy of the covariance matrix $(Y_t, Z_t, U_t, V_t, W_{1t})$ this implies that the cubic term is dominated by the quadratic term, and the right-hand side of (4.6.15) reduces to

$$\begin{aligned}
&2\{\bar t_1\sum_{t=1}^n Y_t+\bar t_2\sum_{t=1}^n Z_t+\bar t_3\sum_{t=1}^n U_t+\bar t_4\sum_{t=1}^n V_t+\bar t_5\sum_{t=1}^n W_{1t}\}\\
&-\{\bar t_1\sum_{t=1}^n Y_t+\bar t_2\sum_{t=1}^n Z_t+\bar t_3\sum_{t=1}^n U_t+\bar t_4\sum_{t=1}^n V_t+\bar t_5\sum_{t=1}^n W_{1t}\}^2\{1+o_P(1)\}\\
&+O_P(\bar\epsilon_n).
\end{aligned}\tag{4.6.16}$$

In a next step, we show that

$$\bar\epsilon_n=o_P(1)+o_P(n)\{\sum_{l=1}^5 \bar t_l^2\},\tag{4.6.17}$$

which is a consequence of the following lemma.

Lemma 4.3. *Under the conditions of Lemma 4.1 and the null model we have for* $h=1,2$,

$$\bar\zeta_h^5=o_P\big(\sum_{l=1}^5|\bar t_l|\big),\quad \bar\zeta_h(\bar\sigma_h^2-1)^2=o_P\big(\sum_{l=1}^5|\bar t_l|\big)\ \textit{and}\ (\bar\sigma_h^2-1)^3=o_P\big(\sum_{l=1}^5|\bar t_l|\big).$$

Proof. Note that by Lemma 4.1 $\bar t_l=o_P(1)$ for $l=1,\dots,5$. Let $\overline\beta_h=\bar\zeta_h^2+(\bar\sigma_h^2-1)$, $h=1,2$, i.e. $\bar t_2=(1-\overline\alpha)\overline\beta_1+\overline\alpha\overline\beta_2$. Due to symmetry, we confine our attention to $h=1$: The results to be shown for $h=1$ hold also true for $h=2$.
By the definition of $\bar t_1$ we have

$$\bar\zeta_2=\{\bar t_1-(1-\overline\alpha)\bar\zeta_1\}/\overline\alpha\tag{4.6.18}$$

and

$$\overline\beta_2=\{\bar t_2-(1-\overline\alpha)\overline{\beta_1}\}/\overline\alpha\tag{4.6.19}$$

by the definition of $\bar t_2$. Putting (4.6.18) and (4.6.19) into the definition of $\bar t_3$ and replacing $\bar\sigma_h^2-1$ by $\overline\beta_h-\bar\zeta_h^2$ therein, we get

$$\begin{aligned}
\bar t_3=&\,3\frac{1-\overline\alpha}{\overline\alpha}\left\{\bar\zeta_1\overline\beta_1-\frac{2(2\overline\alpha-1)}{3\overline\alpha}\bar\zeta_1^3\right\}\\
&+\bar t_1\frac{1}{\overline\alpha^2}\left\{-2\bar t_1^2+3\overline\beta_1\overline\alpha^2-3\overline\beta_1\overline\alpha+6\overline{t_1}\bar\zeta_1(1-\overline\alpha)\right\}\\
&+\bar t_1\frac{6}{\overline\alpha^2}\left\{-\bar\zeta_1^2(1-\overline\alpha)^2\right\}\\
&+\bar t_2\frac{3}{\overline\alpha}\left\{\bar t_1+\bar\zeta_1(\overline\alpha-1)\right\}.
\end{aligned}$$

By Lemma 4.1 we obtain

$$\bar{t}_3 = 3\frac{1-\overline{\alpha}}{\overline{\alpha}}\left\{\overline{\zeta}_1\overline{\beta}_1 - \frac{2(2\overline{\alpha}-1)}{3\overline{\alpha}}\overline{\zeta}_1^3\right\} + o_P(|\bar{t}_1|) + o_P(|\bar{t}_2|). \tag{4.6.20}$$

Plugging (4.6.18) and (4.6.19) into the definiton of $\bar{t}_4$ and replacing $\overline{\sigma}_h^2 - 1$ by $\overline{\beta}_h - \overline{\zeta}_h^2$ therein, it follows that

$$\begin{aligned}\bar{t}_4 =& 3\,\frac{1-\overline{\alpha}}{\overline{\alpha}}\left\{\overline{\beta}_1^2 - \frac{2(1-3\overline{\alpha}+3\overline{\alpha}^2)}{3\overline{\alpha}^2}\overline{\zeta}_1^4\right\}\\ &+ \frac{\bar{t}_1}{\overline{\alpha}^3}\left\{-\bar{t}_1^3 + 8(1-\overline{\alpha})\bar{t}_1^2\overline{\zeta}_1 - 12(1-\overline{\alpha})^2\bar{t}_1\overline{\zeta}_1^2 + 8\overline{\zeta}_1^3(1-\overline{\alpha})^3\right\}\\ &+ 3\bar{t}_2\left\{2\frac{\overline{\alpha}-1}{\overline{\alpha}}\overline{\beta}_1 + \frac{\bar{t}_2}{\overline{\alpha}}\right\}.\end{aligned}$$

By Lemma 4.1 we have

$$\bar{t}_4 = 3\frac{1-\overline{\alpha}}{\overline{\alpha}}\left\{\overline{\beta}_1^2 - \frac{2(1-3\overline{\alpha}+3\overline{\alpha}^2)}{3\overline{\alpha}^2}\overline{\zeta}_1^4\right\} + o_P(|\bar{t}_1|) + o_P(|\bar{t}_2|). \tag{4.6.21}$$

Multiplying (4.6.20) by $\overline{\beta}_1 + \frac{2(2\overline{\alpha}-1)}{3\overline{\alpha}}\overline{\zeta}_1^2$ and (4.6.21) by $\overline{\zeta}_1$ and subtracting, we obtain

$$\left\{\overline{\beta}_1 + \frac{2(2\overline{\alpha}-1)}{3\overline{\alpha}}\overline{\zeta}_1^2\right\}\bar{t}_3 - \overline{\zeta}_1\bar{t}_4 = \frac{2(1-\overline{\alpha})(1-\overline{\alpha}+\overline{\alpha}^2)}{3\overline{\alpha}^3}\overline{\zeta}_1^5 + o_P(|\bar{t}_1|) + o_P(|\bar{t}_2|).$$

Note here that the coefficient of $\bar{t}_3$ in the above equation is $o_P(1)$ since we assume $\overline{\alpha} \in [\delta', 1-\delta']$ and by Lemma 4.1 $\overline{\zeta}_1^2$ and $\overline{\beta}_1$ are $o_P(1)$. The coefficient of $\bar{t}_4$ is $o_P(1)$ since $\overline{\zeta}_1$ is $o_P(1)$ (due to Lemma 4.1). Since the coefficients of $\bar{t}_3$ and $\bar{t}_4$ are $o_P(1)$ and the coefficient of $\overline{\zeta}_1^5$ is bounded away from 0 (since $1-\overline{\alpha}+\overline{\alpha}^2 > 0$ and $1-\overline{\alpha} \geq \delta'$ which is due to the assumption $\overline{\alpha} \in [\delta', 1-\delta']$), we get

$$\overline{\zeta}_1^5 = o_P\Big(\sum_{l=1}^{5}|\bar{t}_l|\Big). \tag{4.6.22}$$

Multiplying (4.6.21) by $\overline{\zeta}_1$ and using $\overline{\zeta}_1^5 = o_P\big(\sum_{l=1}^5|\bar{t}_l|\big)$, we get $\overline{\zeta}_1\overline{\beta}_1^2 = o_P\big(\sum_{l=1}^5|\bar{t}_l|\big)$ since $\overline{\alpha} \in [\delta', 1-\delta']$. Note that

$$|\overline{\zeta}_1(\overline{\sigma}_1^2 - 1)^2| \leq 2|\overline{\zeta}_1|(\overline{\beta}_1^2 + \overline{\zeta}_1^4) = 2|\overline{\zeta}_1\overline{\beta}_1^2| + 2|\overline{\zeta}_1^5|.$$

by the definition of $\overline{\beta}_1$ and by the inequality $(a-b)^2 \leq 2a^2 + 2b^2$, $a, b \geq 0$. By $\overline{\zeta}_1\overline{\beta}_1^2 =$

$o_P\big(\sum_{l=1}^5 |\bar{t}_l|\big)$ and $\overline{\zeta}_1^5 = o_P\big(\sum_{l=1}^5 |\bar{t}_l|\big)$ we get

$$\overline{\zeta}_1(\overline{\sigma}_1^2 - 1)^2 = o_P\big(\sum_{l=1}^5 |\bar{t}_l|\big). \tag{4.6.23}$$

Multiplying (4.6.20) by $\frac{2(1-3\overline{\alpha}+3\overline{\alpha}^2)}{3\overline{\alpha}^2}\overline{\zeta}_1^3$ and (4.6.21) by $\overline{\beta}_1$ and adding yields

$$\frac{2(1-3\overline{\alpha}+3\overline{\alpha}^2)}{3\overline{\alpha}^2}\overline{\zeta}_1^3\bar{t}_3+\overline{\beta}_1\bar{t}_4 = 3\frac{1-\overline{\alpha}}{\overline{\alpha}}\{\overline{\beta}_1^3+\frac{4(1-3\overline{\alpha}+3\overline{\alpha}^2)(2\overline{\alpha}-1)}{9\overline{\alpha}^3}\overline{\zeta}_1^6\}+o_P(|\bar{t}_1|)+o_P(|\bar{t}_2|).$$

Since $\overline{\zeta}_1^6 = o_P(1)\overline{\zeta}_1^5 = o_P\big(\sum_{l=1}^5 |\bar{t}_l|\big)$ by Lemma 4.1 and (4.6.22) this leads to

$$\frac{2(1-3\overline{\alpha}+3\overline{\alpha}^2)}{3\overline{\alpha}^2}\overline{\zeta}_1^3\bar{t}_3+\overline{\beta}_1\bar{t}_4 = 3\frac{1-\overline{\alpha}}{\overline{\alpha}}\overline{\beta}_1^3 + o_P\big(\sum_{l=1}^5 |\bar{t}_l|\big).$$

Given that $(1-3\overline{\alpha}+3\overline{\alpha}^2)/(3\overline{\alpha}^2)$ and $(1-\overline{\alpha})/\overline{\alpha}$ are bounded (by the assumption $\overline{\alpha} \in [\delta', 1-\delta']$),

$$|\overline{\beta}_1|^3 = o_P\big(\sum_{l=1}^5 |\bar{t}_l|\big). \tag{4.6.24}$$

Since $\overline{\beta}_1 = \overline{\zeta}_1^2 + (\overline{\sigma}_1^2 - 1)$ and using the inequality $(a+b)^3 \leq 4(a^3+b^3)$ for $a, b \geq 0$ we have

$$|\overline{\sigma}_1^2 - 1|^3 = |\overline{\beta}_1 + \overline{\zeta}_1^2|^3 \leq 4|\overline{\beta}_1|^3 + 4\overline{\zeta}_1^6 = o_P\big(\sum_{l=1}^5 |\bar{t}_l|\big), \tag{4.6.25}$$

where the last equality is due to (4.6.22) and (4.6.24). □

Lemma 4.1 gives $(\overline{\phi}-\phi_0)^2 = o_P(|\overline{\phi}-\phi_0|) = o_P(|\bar{t}_1|) = o_P\big(\sum_{l=1}^5 |\bar{t}_l|\big)$ and by the same reasoning $|\overline{\zeta}_h||\overline{\phi}-\phi_0| = o_P\big(\sum_{l=1}^5 |\bar{t}_l|\big)$ and $|\overline{\sigma}_h^2 - 1||\overline{\phi}-\phi_0| = o_P\big(\sum_{l=1}^5 |\bar{t}_l|\big)$, $h = 1, 2$. Therefore, we have by Lemma 4.3

$$\overline{\epsilon}_n = n^{1/2} o_P\big(\sum_{l=1}^5 |\bar{t}_l|\big).$$

Using the inequalities $|x| \leq 1 + x^2$ and $(a+b)^2 \leq 2(a^2+b^2)$, $a, b \geq 0$ repeatedly, we get

$$\overline{\epsilon}_n = o_P(1) + n o_P\big(\sum_{l=1}^5 \bar{t}_l^2\big)$$

as claimed. By the non-degeneracy of the covariance matrix $(Y_t, Z_t, U_t, V_t, W_{1t})$ this means that a stochastic upper bound for r_{1n} (which strengthens (4.6.16)) is given by

$$\begin{aligned}
&2\{\bar{t}_1\sum_{t=1}^{n} Y_t + \bar{t}_2\sum_{t=1}^{n} Z_t + \bar{t}_3\sum_{t=1}^{n} U_t + \bar{t}_4\sum_{t=1}^{n} V_t + \bar{t}_5\sum_{t=1}^{n} W_{1t}\} \\
&-\{\bar{t}_1\sum_{t=1}^{n} Y_t + \bar{t}_2\sum_{t=1}^{n} Z_t + \bar{t}_3\sum_{t=1}^{n} U_t + \bar{t}_4\sum_{t=1}^{n} V_t + \bar{t}_5\sum_{t=1}^{n} W_{1t}\}^2\{1+o_P(1)\} \\
&+o_P(1).
\end{aligned} \tag{4.6.26}$$

Note that Y_t, Z_t, U_t, V_t and W_{1t} are mutually orthogonal (see Section 3.6.1). Therefore

$$\begin{aligned}
r_{1n} \le\, & 2\bar{t}_1\sum_{t=1}^{n} Y_t - \bar{t}_1^2\sum_{t=1}^{n} Y_t^2\{1+o_p(1)\} \\
&+2\bar{t}_2\sum_{t=1}^{n} Z_t - \bar{t}_2^2\sum_{t=1}^{n} Z_t^2\{1+o_p(1)\} \\
&+2\bar{t}_3\sum_{t=1}^{n} U_t - \bar{t}_3^2\sum_{t=1}^{n} U_t^2\{1+o_p(1)\} \\
&+2\bar{t}_4\sum_{t=1}^{n} V_t - \bar{t}_4^2\sum_{t=1}^{n} V_t^2\{1+o_p(1)\} \\
&+2\bar{t}_5\sum_{t=1}^{n} W_{1t} - \bar{t}_5^2\sum_{t=1}^{n} W_{1t}^2\{1+o_p(1)\} + o_P(1).
\end{aligned} \tag{4.6.27}$$

Using the properties of quadratic functions we have

$$r_{1n} \quad \le \quad \frac{(\sum_{t=1}^n Y_t)^2}{\sum_{t=1}^n Y_t^2} + \frac{(\sum_{t=1}^n Z_t)^2}{\sum_{t=1}^n Z_t^2} + \frac{(\sum_{t=1}^n U_t)^2}{\sum_{t=1}^n U_t^2} + \frac{(\sum_{t=1}^n V_t)^2}{\sum_{t=1}^n V_t^2} + \frac{(\sum_{t=1}^n W_{1t})^2}{\sum_{t=1}^n W_{1t}^2} + o_P(1).$$

Proof of Theorem 4.1. Since

$$\begin{aligned}
&2\{pl_n(0.5,\widehat{\zeta}_0,\widehat{\zeta}_0,\widehat{\phi}_0,\widehat{\sigma}_0,\widehat{\sigma}_0) - pl_n(0.5,0,0,\phi_0,1,1)\} \\
&= 2\{l_n(0.5,\widehat{\zeta}_0,\widehat{\zeta}_0,\widehat{\phi}_0,\widehat{\sigma}_0,\widehat{\sigma}_0) - l_n(0.5,0,0,\phi_0,1,1)\} \\
&= \frac{(\sum_{t=1}^n Y_t)^2}{\sum_{t=1}^n Y_t^2} + \frac{(\sum_{t=1}^n Z_t)^2}{\sum_{t=1}^n Z_t^2} + \frac{(\sum_{t=1}^n W_{1t})^2}{\sum_{t=1}^n W_{1t}^2} + o_P(1),
\end{aligned} \tag{4.6.28}$$

we get

$$\begin{aligned}
&2\{pl_n(\overline{\alpha},\overline{\zeta}_1,\overline{\zeta}_2,\overline{\phi},\overline{\sigma}_1,\overline{\sigma}_2) - pl_n(0.5,\widehat{\zeta}_0,\widehat{\zeta}_0,\widehat{\phi}_0,\widehat{\sigma}_0,\widehat{\sigma}_0)\} \\
&\qquad \le \frac{(\sum_{t=1}^n U_t)^2}{\sum_{t=1}^n U_t^2} + \frac{(\sum_{t=1}^n V_t)^2}{\sum_{t=1}^n V_t^2} + o_P(1).
\end{aligned} \tag{4.6.29}$$

showing that the χ_2^2 distribution serves as a stochastic upper bound for our test statistic $EM_n^{(K)}$. It remains to show that this upper bound is also attained, asymptotically. Due to the *EM-property* it suffices to find adequate parameters $\alpha, \zeta_1, \zeta_2, \sigma_1, \sigma_2$ and ϕ such that (4.6.29) becomes an equality. This is equivalent to finding a set of values such that $t_j = \hat{t}_j + o_P(n^{-1/2})$, $j = 1, \ldots, 5$, where the t_j's are defined in (4.6.14) and

$$\hat{t}_1 = \frac{\sum_{t=1}^n Y_t}{\sum_{t=1}^n Y_t^2},\ \hat{t}_2 = \frac{\sum_{t=1}^n Z_t}{\sum_{t=1}^n Z_t^2},\ \hat{t}_3 = \frac{\sum_{t=1}^n U_t}{\sum_{t=1}^n U_t^2},\ \hat{t}_4 = \frac{\sum_{t=1}^n V_t}{\sum_{t=1}^n V_t^2},\ \hat{t}_5 = \frac{\sum_{t=1}^n W_{1t}}{\sum_{t=1}^n W_{1t}^2}.$$

Fixing $\alpha = 0.5$ and ignoring terms of order $o_p(n^{-1/2})$ we are searching for parameters $\zeta_1, \zeta_2, \sigma_1, \sigma_2$ and ϕ satisfying

$$\begin{aligned} \hat{t}_1 &= \frac{1}{2}(\zeta_1 + \zeta_2), \\ \hat{t}_2 &= \frac{1}{2}(\beta_1 + \beta_2), \\ \hat{t}_3 &= 3\zeta_1\beta_1, \\ \hat{t}_4 &= 3\beta_1^2 - 2\zeta_1^4, \\ \hat{t}_5 &= (\phi - \phi_0), \end{aligned} \tag{4.6.30}$$

where $\hat{t}_3 = 3\zeta_1\beta_1 + o_P(n^{-1/2})$ is due to (4.6.20) while $\hat{t}_4 = 3\beta_1^2 - 2\zeta_1^4 + o_P(n^{-1/2})$ follows from (4.6.21). Define

$$g(x) = 6x^6 + 3\hat{t}_4 x^2 - \hat{t}_3^2, \quad x \in \mathbb{R}.$$

The third and fourth equation in (4.6.30) imply

$$\begin{aligned} g(\zeta_1) &= 6\zeta_1^6 + 3\hat{t}_4\zeta_1^2 - \hat{t}_3^2 \\ &= 6\zeta_1^6 + 3(3\beta_1^2 - 2\zeta_1^4)\zeta_1^2 - (3\zeta_1\beta_1)^2 \\ &= 6\zeta_1^6 + 9\beta_1^2\zeta_1^2 - 6\zeta_1^6 - 9\zeta_1^2\beta_1^2 \\ &= 0, \end{aligned} \tag{4.6.31}$$

Therefore, we have to find a root of $g(\cdot)$ to obtain an adequate $\tilde{\zeta}_1$. Since $g(0) < 0$ and $g(x) > 0$ for $x \to \infty$, there exists a positive root of $g(\cdot)$. Therefore we can choose $\tilde{\zeta}_1$ to be the smallest positive root of $g(\cdot)$. By (4.6.30) we have $\tilde{\beta}_1 = \hat{t}_3/(3\tilde{\zeta}_1)$, $\tilde{\zeta}_2 = 2\hat{t}_1 - \tilde{\zeta}_1$, $\tilde{\beta}_2 = 2\hat{t}_2 - \tilde{\beta}_1$ and $\tilde{\phi} = \hat{t}_5 - \phi_0$.

By the same reasoning as in Section 3.6, we get $\tilde{t}_j = \hat{t}_j + o_P(n^{-1/2})$ and $\hat{t}_j = O_P(n^{-1/2})$, $j = 1, \ldots, 5$. Putting $\tilde{t}_j$, $j = 1, \ldots 5$, into the right-hand side of (4.6.27) we see that the upper bound of $r_{1n}(\cdot)$ is also attained. Exemplary, we show this for $j = 1$:

$$\begin{aligned}
& 2\tilde{t}_1 \sum_{t=1}^{n} Y_t - \tilde{t}_1^2 \sum_{t=1}^{n} Y_t^2\{1 + o_P(1)\} \\
= \;& 2\big(\hat{t}_1 + o_P(n^{-1/2})\big) \sum_{t=1}^{n} Y_t - \big(\hat{t}_1 + o_P(n^{-1/2})\big)^2 \sum_{t=1}^{n} Y_t^2\{1 + o_P(1)\} \\
= \;& 2\frac{(\sum_{t=1}^{n} Y_t)^2}{\sum_{t=1}^{n} Y_t^2} - \Big\{\left(\frac{\sum_{t=1}^{n} Y_t}{\sum_{t=1}^{n} Y_t^2}\right)^2 + \underbrace{O_P(n^{-1/2})o_P(n^{-1/2})}_{=o_P(n^{-1})} + o_P(n^{-1})\Big\} \sum_{t=1}^{n} Y_t^2\{1 + o_P(1)\} \\
& + o_p(1) \\
= \;& 2\frac{(\sum_{t=1}^{n} Y_t)^2}{\sum_{t=1}^{n} Y_t^2} - \left\{\frac{(\sum_{t=1}^{n} Y_t)^2}{\sum_{t=1}^{n} Y_t^2} + o_P(1)\right\}\{1 + o_P(1)\} + o_p(1) \\
= \;& \frac{(\sum_{t=1}^{n} Y_t)^2}{\sum_{t=1}^{n} Y_t^2} + o_P(1).
\end{aligned}$$

Thereby we used the definition of $\hat{t}_1$, $\hat{t}_1 = O_P(n^{-1/2})$ and the CLT to obtain the second equality. The same argumentation holds true for $j = 2, 3, 4$. To obtain the third equality, we used the SLLN. For $j = 5$ we have to use the CLT for stationary and ergodic martingale differences and the ergodic theorem. Therefore, we have

$$\begin{aligned}
& r_{1n}(0.5, \tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\phi}, \tilde{\sigma}_1, \tilde{\sigma}_2) \\
&= \frac{(\sum_{t=1}^{n} Y_t)^2}{\sum_{t=1}^{n} Y_t^2} + \frac{(\sum_{t=1}^{n} Z_t)^2}{\sum_{t=1}^{n} Z_t^2} + \frac{(\sum_{t=1}^{n} U_t)^2}{\sum_{t=1}^{n} U_t^2} + \frac{(\sum_{t=1}^{n} V_t)^2}{\sum_{t=1}^{n} V_t^2} + \frac{(\sum_{t=1}^{n} W_{1t})^2}{\sum_{t=1}^{n} W_{1t}^2} + o_P(1)
\end{aligned}$$

and hence

$$\begin{aligned}
M_n^{(0)}(0.5) &= 2\{pl_n(0.5, \tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\phi}, \tilde{\sigma}_1, \tilde{\sigma}_2) - pl_n(0.5, \widehat{\zeta}_0, \widehat{\zeta}_0, \widehat{\phi}_0, \widehat{\sigma}_0, \widehat{\sigma}_0)\} \\
&= r_{1n}(0.5, \tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\phi}, \tilde{\sigma}_1, \tilde{\sigma}_2) + r_{2n} \\
&= \frac{(\sum_{t=1}^{n} U_t)^2}{\sum_{t=1}^{n} U_t^2} + \frac{(\sum_{t=1}^{n} V_t)^2}{\sum_{t=1}^{n} V_t^2} + o_P(1).
\end{aligned}$$

By the *EM-property*, we have

$$\begin{aligned}
EM_n^{(K)} \geq M_n^{(0)}(0.5) =\;& 2\{pl_n(0.5, \tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\phi}, \tilde{\sigma}_1, \tilde{\sigma}_2) - pl_n(0.5, \widehat{\zeta}_0, \widehat{\zeta}_0, \widehat{\phi}_0, \widehat{\sigma}_0, \widehat{\sigma}_0)\} \\
=\;& \frac{(\sum_{t=1}^{n} U_t)^2}{\sum_{t=1}^{n} U_t^2} + \frac{(\sum_{t=1}^{n} V_t)^2}{\sum_{t=1}^{n} V_t^2} + o_P(1),
\end{aligned}$$

and therefore the limiting distribution of $EM_n^{(K)}$ is given by the χ_2^2 distribution. □

5 Outlook

In this thesis we were mainly concerned with the basic methodological issue of testing for regime switching in Markov-switching autoregressive models. To this end, we extended the work of Cho and White (2007) using penalized likelihood based tests. We gave the limiting distributions of several tests, e.g. the modified (quasi) likelihood ratio test (MQLRT) for testing the hypothesis of no regime switch in regime-switching ARCH models. Since the GARCH(1,1) model is a benchmark model for modeling log returns of asset prices it would be desirable to extend this test to a wider class of models, the so called regime-switching GARCH models, which capture the nice properties of GARCH models and switching autoregressive models

$$X_t = \sigma_t \epsilon_t; \quad \sigma_t^2 = \vartheta_{S_t} + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2.$$

This model does not match our model specification (1.3.3), since σ_t^2 depends on the whole regime path rather than on S_t, only. This causes a likelihood intractable quickly as soon as the number of observations increases. Following Gray (1996) and Xie and Yu (2005) we consider a so called reduced regime-switching GARCH model

$$X_t = \sigma_t \epsilon_t; \quad \sigma_t^2 = \vartheta_{S_t} + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + E\left[\sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \,\middle|\, \mathcal{S}_{t-1} \right],$$

where $\mathcal{S}_{t-1}$ is the σ-algebra generated by $S_{t-1}, S_{t-2}, \ldots$ which overcomes the problem of path dependence. While theoretical results such as consistency or asymptotic normality for the MLE in this model were obtained by Xie and Yu (2005), simulations indicate that there should be reasonable hope that the MQLRT for testing the hypothesis of no regime switch in e.g. a reduced switching GARCH(1,1) asymptotically admits a mixture of a point mass at zero and a χ_1^2 distribution with equal weights.

Rejecting the hypothesis of one regime we know that there will be at least two regimes. In this case, the determination of the true number of regimes remains an open problem. It would be desirable to find an adequate alternative to the time demanding Bootstrap approach introduced by McLachlan (1987) in the case of mixture models. In a first step, testing for two states in a Markov-switching autoregressive model against (at least) three states could be performed via a MQLRT extending the work of Chen, Chen and Kalbfleisch (2004) for mixture models and of Dannemann and Holzmann (2008) for HMMs, especially in case of model (2.2.3).

Bibliography

Andrews, D.W.K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* **69**, 683–734.

Andrews, L.C. (1986). *Special Functions for Engineers and Applied Mathematicians.* Macmillan Publishing Company, New York.

Bhar, R. and Hamori, S. (2004). *Hidden Markov Models.* Advanced Studies in Theoretical and Applied Econometrics. Kluwer Academic Publishers, Dortrecht.

Billingsley, P. (1995). *Probability and measure.* Wiley, New York.

Boshnakov, G.N. (2006). Prediction with mixture autoregressive models. Research Report No. 6, 2006, Probability and Statistics Group School of Mathematics, The University of Manchester.

Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* **52**, 115–127.

Bradley, R.C. (2005). Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions., *Probability Surveys* **2**, 107–144

Brockwell, P.J. and Davis, R.A. (2006). *Time series: Theory and methods.* Springer, New York.

Cai, J. (1994). A Markov model of unconditional variance in ARCH. *Journal of Business and Economic Statistics* **12**, 309–316.

Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in Hidden Markov Models.* Springer, New York.

Chen, J. (1995). Optimal rate of convergence for finite mixture models. *Ann. Stat.* **23**, 221-233.

Chen, H., Chen, J. and Kalbfleisch, J.D. (2001). A modified likelihood ratio test for homogeneity in finite mixture models. *J. R. Stat. Soc. Ser. B* **63**, 19–29.

Chen, H. and Chen, J. (2003). Tests for homogeneity in normal mixtures in the presence of a structural parameter. *Statistica Sinica* **13**, 351-365.

Chen, H., Chen, J. and Kalbfleisch, J. D. (2004). Testing for a finite mixture model with two components. *J. R. Stat. Soc. Ser. B* **66**, 95–115.

Chen, J. and Li, P. (2009). Hypothesis test for Normal Mixture Models: the EM Approach. *Annals of Statistics* **37**, 2523–2542.

Cheng, R.C.H. and Traylor, L. (1995). Non-regular Maximum Likelihood Problems. *J. R. Statist. Soc. B* **57**. 3–44.

Cho, J.S. and White, H. (2007). Testing for regime switching. *Econometrica* **75**, 1671–1720.

Dacunha-Castelle, D. and Gassiat, E. (1999). Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes. *Ann. Statist.* **27**, 1178–1209.

Dannemann, J. (2009, published 2010). *Inference for Hidden Markov Models and related Models*. Ph.D. thesis, University of Göttingen. Cuvillier Verlag, Göttingen.

Dannemann, J. and Holzmann, H. (2008). Testing for two states in a hidden Markov model. *Canadian J. Statist* **36**, 505–520.

Davig, T. (2004). Regime-Switching Debt and Taxation. *Journal of Monetary Economics* **51**, 837–859.

Dempster, A. P., Laird N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via EM algorithm (with discussion). *J.R. Stat. Soc. Ser. B* **39**, 1–38.

Douc, R., Moulines, E. and Rydén T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Annals of Statistics* **32**, 2254–2304.

Dupont, R. (2010). Modellwahlkriterien für Switching-Regressions-Modelle. Diploma thesis, KIT Karlsruhe. (in German)

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York.

Ferguson, T.S. (1996). *A Course in Large Sample Theory*. Chapman & Hall, London.

Francq C. and Zakoïan J.M. (2001). Stationarity of multivariate Markov–switching ARMA models. *Journal of Econometrics* **102**, 339–364

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer, New York.

Frydman, R. (1980). A proof of the consistency of maximum likelihood estimators of nonlinear regression models with autocorrelated errors. *Econometrica* **48**, 853–860.

Fu, Y., Chen, J. and Li, P. (2008). Modified likelihood ratio test for homogeneity in a mixture of von Mises distributions. *Journal of Statistical Planning and Inference* **138**, 667–681.

Gassiat, E. and Keribin, C. (2000). The likelihood ratio test for the number of components in a mixture with Markov regime. *ESAIM P&S* **4**, 25–52.

Gassiat, E. (2000). Likelihood ratio inequalities with applications to various mixtures.

Ann. Inst.Henri Poincaré **38**, 897–906.

Goffinet, B., Loisel, P. and Laurent, B. (1992). Testing in normal mixture models when the proportions are known. *Biometrika* **79**, 842–846.

Gray, S.F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* **42**, 27–62.

Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57**, 357–384.

Hamilton J.D. and Susmel R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* **64**, 307–333.

Hamilton, J.D. (2008). Regime-Switching Models. In *The New Palgrave Dictionary of Economics, Second Edition* (eds. S.N. Durlauf and L.E. Blume). Palgrave McMillan Ltd.

Hathaway, R.J. (1985). A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *Annals of Statistics* **13**, 795–800.

Henze, N. (2002). Invariant tests for multivariate normality: A critical review. *Statistical Papers* **43**, 467–506.

Holzmann, H., Munk, A. and Gneiting, T. (2006). Identifiability of finite mixtures of elliptical distributions. *Scandinavian J. Statist* **33**, 753–763.

Hosmer, D.W. (1973). A comparision of iterative maximum-likelihood estimates of the parameters of a mixture of two normal distributions under three different types of sample. *Biometrics* **29**, 761–770.

Hurst, S. (1995). The characteristic function of student t distribution. Financial Mathematics Research Report No. FMRR 006-95, Statistical Research Report No. SRR044-95.

Juang, B.H. , Rabiner, L.R. (1985). Mixture autoregressive hidden Markov models for speech signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **33**, 1404–1413.

Kim, C.J., Nelson, C.R. and Startz, R. (1998). Testing for mean reversion in heteroscedastic data based on Gibbs-sampling augmented randomization. *Journal of Empirical Finance* **5**, 131–154.

Kon, S.J. (1984). Models of stock returns – A comparison. *The Journal of Finance* **39**, 147–165.

Kreiss, J.P. and Neuhaus, G. (2006). *Einführung in die Zeitreihenanalyse*. Springer, Berlin. (in German)

Krengel, U. (1985). *Ergodic Theorems*. De Gruyter Studies in Mathematics, Berlin.

Krolzig, H.M. (1997). *Markov-switching vector autoregressions*. Lecture notes in Economics and Mathematical Systems 454. Springer, Berlin.

Lange, T. and Rahbek A. (2009). An Introduction to Regime Switching Time Series Models. In *Handbook of Financial Time Series* (R.A. Davis, J.P. Kreiss and T. Mikosch,eds.). Springer, New York.

Leroux, B.G. (1992a). Consistent estimation of a mixing distribution. *Annals of Statistics* **20**, 1350–1360.

Leroux, B.G. (1992b). Maximum-likelihood estimation for hidden Markov models. *Stochastic Process. Appl.* **40**, 127–143.

Li, P., Chen, J. and Marriot, P. (2009). Non-finite Fisher information and homogeneity: the EM approach. *Biometrika* **96**, 411–426.

Liu, X. and Shao, Y. (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *Ann. Statist.* **31**, 807–832.

Lu, Z. (1998). On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term. *Statistica Sinica* **8**, 1205–1217.

McCulloch, R.E. and Tsay, R.S. (1994). Statistical inference of macroeconomic time series via Markov switching models. *Journal of Time Series Analysis* **15**, 523–539.

McLachlan, G.J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Journal of the Royal Statistical Society Series C* **36**, 318–324.

McLachlan, G.J. and Peel D. (2000). *Finite Mixture Models.* Wiley Series in Probability and Statistics. New York.

Meng, X.L. and Rubin, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80**, 267–278.

Naik, P.A., Shi, P. and Tsai, C.-L. (2007). Extending the akaike information criterion to mixture regression models. *Journal of the American Statistical Association*, 244–256.

Olteanu, M. and Rynkiewicz, J. (2007). Estimating the number of regimes in a switching autoregressive model. *Working paper*

Pierce, D.A. (1985). Testing normality in autoregressive models. *Biometrika* **72**, 293–297.

Piger, J. (2009). Econometrics: models of regime changes. In *Encyclopedia of Complexity and Systems Science* (R.A. Meyers, ed.). Springer, New York.

Porter, R.H. (1983). A study of cartel stability: The Joint Executive Committee, 1880-1886. *The Bell Journal of Economics* **14**, 301–314.

Poskitt, D.S. and Zhang, J. (2005). Estimating components in finite mixtures and hidden Markov models. *Australian & New Zealand Journal of Statistics* **47**, 269–286.

Potter, S.M. (1999). Nonlinear Time Series Modelling: An Introduction. *Journal of Economic Surveys* **13**, 505–528.

Qin, Y.S. and Smith, B. (2004). Likelihood ratio test for homogeneity in normal mixtures in the presence of a structural parameter. *Statistica Sinica* **14**, 1165–1177.

Sclove, S.L. (1983). Time series segmentation: A model and a method. *Information Science*, **29**, 7–25.

Seidel, W. (2010) Mixture Models. In *International Encyclopedia of Statistical Science* (M. Lovric, ed.). Springer, Heidelberg.

Stelzer, R. (2005). On Markov-Switching Models – Stationarity and Tail Behaviour. Diploma thesis, TU München.

Stelzer, R. (2009). On Markov-Switching ARMA Processes – Stationarity, Existence of Moments and Geometric Ergodicity. *Econometric Theory* **25**, 43–62.

Shumway, R.H. and Stoffer, D.S. (2006). *Time Series Analysis and Its Applications. With R Examples.* Springer, New York.

Teicher, H. (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics* **34**, 1265–1269.

Teicher, H. (1967). Identifiability of mixtures of product measures. *The Annals of Mathematical Statistics* **38**, 1300-1302.

Titterington, D.M., Smith A.F.M. and Makov U.E. (1985). *Statistical Analysis of Finite Mixture Distributions.* Wiley Series in Probability and Statistics. New York.

Tong, H. (1983). *Threshold models in non-linear time series analysis.* Springer, New York.

Tsay, R.S. (2002). *Analysis of financial time series.* John Wiley and Sons, New York.

Velucchi, M. (2009). Regime Switching: Italian financial markets over a century. *Stat. Methods Appl.*, 67–86.

Viterbi, A.J. (1967). Error bounds for convolutional codes and an asymptotically optimal decoding algorithm. *IEEE TRans. Inorm. Th.* **13**, 260–269.

Vollmer, S., Holzmann, H., Ketterer, F. and Klasen S. (2010). Distribution Dynamics of Regional GDP per Employee in Unified Germany. *Preprint.*

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics* **20**, 595–601.

Wong, C.S. and Li, W.K. (2000). On a mixture autoregressive model. *J. Royal Statist. Soc. B* **62**, 95–115.

Wong, C.S. and Li, W.K. (2001). On a mixture autoregressive conditional heteroscedastic model. *J. American Statist. Assoc.* **96**, 982–995.

Xie, Y. and Yu, J. (2005). Consistency of Maximum Likelihood Estimators for the Reduced Regime-Switching GARCH Model. Research Report, Centre of Biostochastics

Swedish University of Agricultural Sciences, Umeå Sweden.

Yakowitz, S.J. and Spragins, J.D. (1968). On the identifiability of finite mixtures. *The Annals of Mathematical Statistics* **39**, 209–214.

Zhang, J. and Stine, R.A. (2001). Autocovariance structure of Markov switching models and model selection. *Journal of Time Series Analysis* **22**, 107–124.

Zucchini, W. and MacDonald I.A. (2010). *Hidden Markov models for time series*. Chapman&Hall.

Zusammenfassung

Eine Vielzahl von Zeitreihenmodellen, wie z.B. lineare autoregressive oder ARCH-Modelle, wird verwendet, um das Verhalten von ökonomischen und Finanzzeitreihen zu analysieren. Da sich jedoch das Verhalten von Zeitreihen über die Zeit häufig ändert, beschreiben solche Zeitreihenmodelle die Daten möglicherweise nicht zufriedenstellend.

Das Markov-Switching-Modell von Hamilton (1989) ist eines der bekanntesten Regime-Switching-Modelle. Dieses Modell bildet mehrere Strukturen ab, die das Verhalten der Zeitreihe in verschiedenen Zuständen charakterisieren. Während der switchende Mechanismus im ursprünglichen Markov-Switching-Modell in ein lineares autoregressives Modell aufgenommen wurde, untersuchten Cai (1994) und Hamilton und Susmel (1994) verschiedene ARCH-Modelle mit Markov-Switching. In Markov-Switching-Modellen wird der switchende Mechansimus durch eine latente Variable, die einer Markov-Kette der Ordnung 1 folgt, gesteuert. Deshalb ist u.a. die Bestimmung der Anzahl der Zustände der versteckten Markov-Kette von großer Bedeutung. In dieser Arbeit beschäftigen wir uns insbesondere mit der grundlegenden methodischen Fragestellung des Testens auf Regime-Switching in diversen Markov-Switching autoregressiven Modellen. Da unter der Hypothese Parameter des vollen Modells nicht identifizierbar sind, ist die asymptotische Verteilung des entsprechenden Likelihood-Quotienten-Tests nicht standard. Dieses Problem tritt auch schon in der eng verwandten Fragestellung des Testens auf Homogenität in Zwei-Komponenten-Mischungsmodellen auf. Um das Problem der Nicht-Identifizierbarkeit zu lösen, entwickelten Chen, Chen und Kalbfleisch (2001) einen penalisierten Likelihood-Quotienten-Test und zeigten, dass die entsprechende Teststatistik eine einfache asymptotische Verteilung hat. Zusätzliche Schwierigkeiten treten auf, wenn wir die Markov-Abhängigkeitsstruktur in die Teststatistik aufnehmen. Deshalb schlugen Cho und White (2007) einen quasi Likelihood-Quotienten-Test auf Regime Switching in autoregressiven Modellen vor, der die Abhängigkeitsstruktur in der versteckten Markov-Kette unter der Alternative vernachlässigt. Wir erweitern diese Vorgehensweise, indem wir penalisierte Likelihood-basierte Tests entwickeln, um Tests mit einfachen asymptotischen Verteilungen zu erhalten.

In Kapitel 1 stellen wir Markov-Switching autoregressive und eng verwandte Modelle vor und diskutieren die Methodik, die wir im Folgenden benutzen.

Der modifizierte Likelihood-Quotienten-Test, der von Chen, Chen und Kalbfleisch (2001) eingeführt wurde, ist eine etablierte Methode zum Testen auf Homogenität in endlichen Mischungsmodellen. In Kapitel 2 erweitern wir diesen Test auf Markov-Switching autoregressive Modelle mit univariatem switchenden Parameter, die gewisse Regularitätsbeding-

ungen erfüllen. Diese Bedingungen sind z.B. für lineare switching autoregressive Modelle mit switchendem Skalenparameter und normalverteilten Innovationen erfüllt. Wir zeigen, dass die asymptotische Verteilung des modifizierten quasi Likelihood-Quotienten-Tests unter der Hypothese eine Mischung aus einer Punktmasse in 0 und einer χ_1^2-Verteilung ist. Schließlich führen wir einen verwandten Test, den sog. EM-Test, ein, der die gleiche asymptotische Verteilung wie der modifizierte quasi Likelihood-Quotienten-Test aufweist.

Für Anwendungen ist das lineare switching autoregressive Modell mit switchendem Intercept und normalverteilten Innovationen von großer Bedeutung, siehe Hamilton (2008). In Kapitel 3 geben wir einen Test auf Homogenität in diesem Modell an. Allerdings ist das Studium asymptotischer Eigenschaften von Teststatistiken, die auf der (penalisierten) Likelihood basieren, sehr schwierig, da $\sigma\frac{\partial^2 f(x;\mu,\sigma)}{\partial^2\mu} = \frac{\partial f(x;\mu,\sigma)}{\partial\sigma}$ für die Normalverteilung gilt. Hierbei ist $f(x;\mu,\sigma)$ die Dichte der Normalverteilung mit Erwartungswert μ und Standardabweichung $\sigma > 0$. Dieses Problem tritt bereits beim Testen auf Homogenität in homoskedastischen Mischungen univariater Normalverteilungen auf. Chen und Li (2009) entwickelten hierfür einen Test. In Kapitel 3 übertragen wir diesen Test auf lineare switching autoregressive Modelle mit normalverteilten Innovationen, in denen der Intercept gemäß des zugrundeliegenden Regimes switcht. Wir zeigen, dass die asymptotische Verteilung der entsprechenden Teststatistik unter der Hypothese eine einfache Funktion der verschobenen χ_1^2- und $\frac{1}{2}\chi_0^2+\frac{1}{2}\chi_1^2$-Verteilung ist. Ferner schlagen wir einen Test gegen feste Gewichte unter der Alternative vor und berechnen die asymptotische Verteilung dieser Teststatistik unter der Hypothese. Wir wenden die in Kapitel 2 und 3 entwickelten Methoden auf die saisonal bereinigten Quartalsdaten des U.S. BIPs von 1947(1) bis 2002(3) an und finden einen Regime Switch in der Volatilität der Wachstumsrate. Schließlich teilen wir die Zeitreihe in zwei Teilzeitreihen 1947(1)-1984(1) und 1984(2)-2002(3) auf und finden keine klare Evidenz für einen Regime Switch im Intercept in einem linearen autoregressiven Modell in diesen Teilreihen.

In Kapitel 4 beschäftigen wir uns mit Tests auf Homogenität in einem linearen switching autoregressiven Modell, in dem sowohl Intercept als auch Varianz der normalverteilten Innnovationen switchen dürfen. Wir erweitern den sog. EM-Test von Chen und Li (2009) zum Testen auf Homogenität in einem Normalverteilungsmischungsmodell mit verschiedenen Erwartungswerten und Varianzen unter der Alternative. Wir zeigen, dass die asymptotische Verteilung der Teststatistik unter der Hypothese eine χ_2^2-Verteilung ist. Da der EM-Test die gleiche asymptotische Verteilung aufweist, wenn wir unter der Alternative das Gewicht $\alpha = 1/2$ festhalten, schlagen wir auch einen Test gegen festes Gewicht $\alpha = 1/2$ unter der Alternative vor. Wir wenden unsere Methoden auf die monatlichen log-Returns der IBM-Aktie an und finden Evidenz für 2 Zustände: Regime 1 mit niedrigerem Intercept und höherer Varianz und Regime 2 mit höherem Intercept und niedriger Varianz.

Curriculum Vitae

Florian Ketterer
born 25th February 1982 in Achern, Germany
Single, German

09/1988–06/2001	Schooling *Abitur* at *Heimschule Lender Sasbach (Gymnasium)*
07/2001–04/2002	Civilian service
10/2002–06/2008	Diploma studies in business mathematics Faculty of Mathematics, University of Karlsruhe Diploma thesis: *Ein graphisches Verfahren zum Vergleich von Normalverteilungstests* Supervisor: *Prof. Dr. Norbert Henze*
07/2008–12/2008	Ph.D. studies in mathematics *Institute for Stochastics, University of Karlsruhe*
since 01/2009	Ph.D. studies in mathematics *Institute for Stochastics and Mathematical Statistics,* *University of Marburg* Supervisor: *Prof. Dr. Hajo Holzmann*